TITLE: System for Electronically Managing, Finding, and/or Displaying Biomolecular Interactions

FIELD OF THE INVENTION

The invention relates to a system, methods and products for managing, finding, and/or displaying biomolecular interactions.

BACKGROUND OF THE INVENTION

Technological advances and mounting interest have pushed proteomics into the scientific spotlight. This growing field encompasses the study of proteins, both in structure and in function, contained in a proteome - the protein equivalent of a genome. Because of increased interest and technique automation (Mendelsohn et al., 1999), the rate of proteomic data production is growing in a similar fashion as that of genomics a decade ago. For example, mass spectrometers, gene chips, and two-hybrid systems have made cellular signaling pathway mapping faster and easier and consequently these are becoming large producers of data. Protein-protein interaction and more general biomolecule-biomolecule (protein-DNA, protein-RNA, protein-small molecule, etc.) interaction information is being generated and recorded in the literature. Lessons from the genomic era have taught us that large amounts of related data recorded in scientific journals soon becomes unmanageable. A well designed common data specification based on a model of the biological information is therefore required to describe and store biomolecular interaction data.

SUMMARY OF THE INVENTION

The present inventors have designed a data specification for the storage and management of biomolecular interaction and biochemical pathway data that possesses the following properties:

1. It describes the full complexity of the biological data, from simple binary interactions to large-scale molecular complexes and networks of pathways and interactions. It stores protein, DNA, RNA, and other molecules in full atomic detail, since character based sequence abstractions of biomolecules often miss important chemical features, such as methylation on DNA. This allows as much data as possible to be stored for scientific use in electronic form rather than in print.

2. It is easily computable. A computer can easily read, write, and traverse the specification. This facilitates maintenance of a database of such information, creation of advanced queries and querying tools and development of computer programs that use the information for data visualization, data mining, and visual data entry.

3. It is platform and database independent. Tools written for one platform can read data created on another platform directly. It handles the data structure without modification as well.

4. It is succinct and easy for humans to understand. Field to data correspondence is very clear and a human readable format of the specification is available.

The data structure was designed for a database referred to herein as "BIND" (Biomolecular Interaction Network Database). The data structure is written in a data specification language called Abstract Syntax Notation.1 (ASN.1, also known as X.208 or ISO-8824) (http://www.oss.com/asn1/index.html). The U.S. National Center for Biotechnology Information (NCBI) uses ASN.1 to describe and store all of its biological and publication data and all of GenBank, MMDB and PubMed (Ostell and Kans, 1998). BIND inherits the NCBI data model, which provides a solid foundation for the BIND data specification through the use of mature NCBI data types that describe sequence, 3D structure, and publication reference information.

Although the specification is written in ASN.1, it is not restricted to this syntax. The data structures can be readily translated to other common data specification languages such as CORBA IDL (Object Management Group, 1996) or XML (http://www.w3.org/XML) if the need arises. Aside from ASN.1, no other biological data specification is sufficiently rich in mature data types to use as a foundation for BIND without first building and testing those base data types.

The BIND data specification represents complex cellular pathway information efficiently in a computer. BIND defines three main data types: interactions, molecular complexes, and pathways. Each of these objects is composed of various component and descriptor objects that are either defined in the specification proper or inherited from the NCBI ASN.1 data specifications. For example, an interaction record contains, among other data objects, two BIND-objects. A BIND-object describes a molecule of any type and is itself defined using simpler sub-objects. Normally, a BIND-object describing a biopolymer sequence will store a simple link to a sequence database, such as GenBank (Benson et al., 1999). If, however, the sequence is not present in a public database, it can be fully represented using an embedded NCBI-Bioseq object. The NCBI-Bioseq object is how NCBI stores all of the sequences in GenBank and is a mature data structure. BIND also inherits the NCBI taxonomy model (also used and supported by EMBL, DDBJ and Swiss-Prot) and data, via an inherited NCBI-BioSource, and is designed so that interactions can be both inter- and intra-organismal. Sequence, structure, publication, taxonomy and small molecule databases provide a strong foundation for BIND.

Broadly stated, the present invention contemplates a system for electronically managing, finding, and/or visualizing biomolecular interactions comprising a computer system including at least one computer receiving data on biomolecular interactions from a plurality of providers and processing such data to create and maintain images and/or text defining biomolecular interactions, said computer system, in response to data requests, creating and transmitting to a plurality of end-users, the images and/or text defining biomolecular interactions.

In an embodiment, a system for electronically managing, finding, and/or visualizing biomolecular interactions is provided comprising:

(a) a maintenance entity for receiving data on biomolecular interactions from a plurality of providers and means for receiving and processing such data to create and maintain images and/or text defining biomolecular interactions; and

(b) one or more computer systems maintained by the maintenance entity and having means for creating and transmitting to a plurality of end-users the images and/or text defining biomolecular interactions.

The system is useful in managing, finding, and/or displaying biomolecular interactions including interactions involving proteins, nucleic acids (RNA, DNA), and ligands, molecular complexes, and signaling pathways. The interactions are defined both at the molecular and atomic levels and in particular they may be defined by chemical graphs.

The invention also provides a method for displaying on a computer screen information concerning biomolecular interactions comprising retrieving an image and/or text defining a biomolecular interaction from a system of the invention.

The present invention also provides a data structure stored in the memory of a computer the data structure having a plurality of records and each record containing a biomolecular interaction and information relating to the biomolecular interaction. In an embodiment the biomolecular interaction is identified by chemical graphs. The information in the data structure may be accessible by using indices which may represent selections of information from the chemical graphs.

The term "record" used herein generally refers to a row in a database table. Each record contains one or more fields or attributes. A given record may be uniquely specified by one or a combination of fields or attributes known as the record's primary key. A record of a biomolecular interaction as used herein is generally a record containing information identifying the biomolecular interaction as a chemical graph and a plurality of other attributes with information pertaining to the biomolecular interaction (e.g. information on the cellular place of interaction, experimental conditions used to observe the interaction, conserved sequence comment of molecules in the interaction if they are biological sequences, information on molecules in the interaction, description of metabolic and signaling pathways, cell cycle stages in which an interaction is involved, locations of binding sites on the molecules in an interaction, chemical actions mediated by the interactions, and chemical states of the molecules in the interaction).

The term "chemical graph" refers to a connectivity graph of all the atoms and bonds in a molecule in a biomolecular interaction. The graph may include three-dimensional coordinates.

The invention also provides a method for storing a representation of a biomolecular interaction in a memory of a computer system, the method executed on a computer system and comprising the steps of:

(a) identifying a chemical graph of a biomolecular interaction; and

(b) storing a record in a data structure of the invention.

The invention further provides a method for storing a representation of a biomolecular interaction in a memory of a computer system, the method executed on a computer system and comprising the steps of:

(a) identifying a chemical graph of a biomolecular interaction;

(b) generating one or more indices from information in the chemical graph; and

(c) storing a record in a data structure of the invention.

The invention still further provides a method for identifying a biomolecular interaction that is similar to a reference biomolecular interaction, the method executed on a computer and comprising the steps of:

(a) conducting a similarity search for each molecule in a test biomolecular interaction;

(b) screening the results of the similarity search preferably by selected taxonomy;

(c) assembling a putative biomolecular interaction to create a test record;

(d) accessing one or more records in a data structure stored in the memory, the data structure having a plurality of records, each of the records containing a reference biomolecular interaction and information relating to the reference biomolecular interaction; and

(e) matching the test record with each record in the data structure to produce a matching record containing a reference biomolecular interaction matching the test biomolecular interaction.

The similarity searches may be based for example on sequence similarity or identity, or similarities in molecular weights, pIs, mass fingerprinting data or mass spectrometric data, fragment-ion tag data, peptide masses from enzymatic digestion, fragment ion masses, isotope patterns, and sequence tag data. Standard tools available in the art for similarity searching and screening can be used. (For example, the following tools may be used BLAST http://www.ncbi.nlm.nih.gov/BLAST/, BioScan, Fasta3, PropSearch, SAMBA, SAWTED, Scanps, FDF, ExPASY Proteomics Tools – http:www.expasy.ch/tools, TagIdent: http://www.expasy.ch/tools/tagident.html, PeptIdent: http://www.expasy.ch/tools/peptident.html, ProteinProspector: http://prospector.ucsf.edu/, MultiIdent: PeptideSearch: http://www.mann.embl-heidelberg.de/Services/PeptideSearch/PeptideSearchIntro.html, PROWL:http://prowl.rockefeller.edu/;Mascot:http://www.matrixscience.com/cgi/index.pl?page=/search_form_select.html; BioSCAN, Pro).

Another aspect of the invention provides a computer system for storing a representation of one or more biomolecular interactions in a memory in the computer system and for comparing one or more reference biomolecular interactions to a test biomolecular interaction, comprising:

(a) a database means stored in the memory representing one or more biomolecular interactions; each of the biomolecular interactions represented by a chemical graph; and

(b) a data structure means for storing a plurality of record means, each record means containing chemical graphs of the test biomolecular interaction.

The invention also provides a computer system comprising memory means, storage means, program means, and stored means for building virtual-models of biomolecular interactions in the computer system comprising:

(a) one or more libraries of reference biomolecular interactions that comprise any number of attributes or components of the biomolecular interaction which values are either being used to describe characteristics of the types of biomiolecular interactions in the computer system, or values or data structures used by the program at runtime, or are to be used to more specifically describe characteristics of individual components of biomolecular

interaction that each instance of a type of biomolecular interaction is to represent, or characteristics of each instance of biomolecular interaction in the computer system; wherein the attributes have values of any type in the computer system or in a network accessible by the computer system;

(b) means for manipulating the biomolecular interaction by domain experts or program means comprising visual means for making the biomolecular interactions available through menus or palettes or programmatic means; and

(c) constructor means to create new instances from the definitions of the biomolecular interactions, and means to establish directional output-input links between complemenatary instances of the biomolecular interactions directly or through components.

Also provided is a computer system comprising:

(a) a database having a plurality of records, wherein each record contains a reference biomolecular interaction defined by a chemical graph and descriptive information from an external database which information correlates the biomolecular interactions to records in the external database; and

(b) a user interface allowing a user to selectively view information regarding a biomolecular interaction.

In an embodiment, a computer system is provided comprising:

(a) a database having a plurality of records, each of said records containing a reference biomolecular interaction defined by a chemical graph and descriptive information from an external database, which information correlates the biomolecular interactions to records in the external database;

(b) a processor in communication with said database and responsive to user input to access records in said database; and

(c) a user interface allowing a user to provide user input to said processor to selectively view information regarding a biomolecular interaction.

Still further the invention provides a database system comprising a plurality of internal records, the database comprising a plurality of records, wherein each record contains a biomolecular interaction defined by chemical graphs and descriptive information from an external database which information correlates the biomolecular interactions to records in the external database.

In an embodiment the external database is PubMed. The interface of the computer system may further comprise user selectable links to enable a user to access additional information for a biomolecular interaction. The links may comprise HTML links.

Additionally provided is a method of using a computer system to present information, or a method of presenting information pertaining to records of biomolecular interactions in a database, the records containing information identifying the biomolecular interaction and defining the biomolecular interaction by chemical graphs, the method comprising:

(a) providing an interface for entering query information relating to a biomolecular interaction;

(b) locating data corresponding to the entered query information; and

(c) displaying the data corresponding to the entered query information.

In step (b) the data is located by examining records in the database.

The invention further provides a computer program product comprising a computer-usable medium having computer-readable program code embodied thereon relating to a plurality of records of biomolecular interactions, the records identifying the biomolecular interactions and defining chemical graphs of the biomolecular interactions, the computer program product comprising computer-readable program code for effecting the following steps within a computing system:

(a) providing an interface for entering query information relating to a biomolecular interaction;

(b) locating data corresponding to the entered query information; and

(c) displaying the data corresponding to the entered query information.

The invention contemplates a database storing data relating to biomolecular interactions comprising:

(a) first data types describing biomolecular interactions between chemical objects;

(b) second data types describing collections of biomolecular interactions; and

(c) third data types describing pathways between said collections of interactions.

The first data types may include objects for the chemical objects, each of the objects including at least one of a pointer to an external database describing the chemical object, a sequence, and a chemical graph. The first data types may be stored as records and further include objects identifying the biomolecular interactions and defining chemical graphs of the biomolecular interactions.

The second data types may include lists of identifications referencing the biomolecular interactions in the collections. The third data types may include objects for the chemical objects that can form networks of interactions. The networks of interactions may include metabolic pathways and cell signaling pathways. The third data types may additionally include sequences of identifications referencing biomolecular interactions that make up the pathways.

The systems and products of the present invention may be used to study and identify biomolecular interactions. Such information is of significant interest in pharmaceutical research, particularly to identify potential drugs and targets for drug development. The systems and products provide great power and flexibility in analyzing biomolecular interactions.

Further features and advantages of the present invention, as well as the structure and operation of various embodiments of the present invention, are described in detail below with reference to the accompany drawings.

DESCRIPTION OF THE DRAWINGS

The invention will be better understood with reference to the drawings in which:

Figure 1 - Storing a chemical object - A BIND-object data type. A chemical object can be any molecule or atom. Associated data types are also shown. Legend: Each box is a data type. Dashed

outline boxes represent ASN.1 fields marked as OPTIONAL. Single headed arrows point to expanded definition for a data type. Double headed arrows represent one to many relationships (repeated fields or objects).

Figure 2 - Storing a chemical object (continued).

Figure 3 - Storing a biomolecular interaction - A BIND-Interaction data structure.

Figure 4 - Storing the cellular place information - A BIND-place object and associated data types. General place is saved using enumerated fields for computability and specific place is more detailed and human-readable.

Figure 5 - Storing experimental condition information - A BIND-condition data object and associated data types.

Figure 6 - Storing conserved sequence information - a BIND-conserved-seq object and associated data types. Conserved sequence may be stored for molecule 'a' or 'b'.

Figure 7 - Storing binding site location - A BIND-loc object and associated data types. Any number of binding sites may be stored for either molecule 'a' or 'b' in an interaction.

Figure 8 - Storing chemical actions - A BIND-action object and associated components. Any number of chemical actions may be stored in an interaction.

Figure 9 - Storing chemical state - A BIND-state data type and associated objects. Any number of chemical states may be stored in an interaction.

Figure 10 - Representing molecular complexes - A BIND-Molecular-Complex object and related data types.

Figure 11 - Storing biochemical pathways - A BIND-Pathway object and associated data types. Cell cycle information can be stored.

Figure 12 is a schematic diagram showing a software development method;

Figure 13 is a schematic diagram showing a major subsystem overview of BIND; and

Figure 14 is a schematic diagram showing the data entry process for BIND.

DESCRIPTION OF PREFERRED EMBODIMENTS OF THE INVENTION

The present inventors have developed BIND or the Biomolecular Interaction Networks Database and its related tools for both the management and mining of molecular interaction data. BIND permits the rapid identification and visualization of new and known cellular pathways using bioinformatics methods, and it provides an understanding of these interaction pathways. BIND is stored in memory within a host computer system including one more computers that is responsible for maintaining BIND. Biomolecular interactions are received from internal and external providers through connections to the host computer network and are processed to maintain images and/or text defining the biomolecular interactions. Images and/or text defining biomolecular interactions in BIND can be conveyed to end-users through computer connections to the host computer system allowing end-users to display the biomolecular interaction data on the monitor display screens of their computers. (See Figure 12 for a schematic diagram of the BIND software development method; Figure 13 for a schematic diagram showing the major components of the BIND system; and Figure 14 for a schematic

diagram of a data entry process for BIND). In this way, biomolecular interactions can be electronically managed, located and/or visualized. Further details concerning BIND will now be described.

Biomolecular information is stored in records within BIND. A BIND record can describe any molecular interaction, stored in a BIND-object as a) a pointer to another database, b) a sequence or c) a chemical graph. Two BIND-objects that interact are held in an interaction record within BIND. The interaction record can represent the binding interaction at various levels of detail. BIND also stores kinetic information, bibliographic information, interaction locations, conserved sequences, mediating interactions, chemical reactions that take place and activation states of BIND-objects. BIND draws upon NCBI data format standards, and thus BIND is compatible with other public sequence and structure databases. BIND forms a data-space that can contain large molecular interactions, such as a protein signaling complex, to detailed descriptions of atomic level interactions. The design and database samples and tools to aid in web-based data entry and retrieval are described herein. A graphical system of data retrieval and data mining agents that scour this data space for novel links between known interaction pathways can be implemented based on this design.

The BIND Data Model

The three main types of data objects in the BIND specification - interaction, molecular complex and pathway - as well as useful database management and data exchange objects are described below. Each of the main objects is composed of various descriptor objects that are either defined in the specification or taken from the NCBI ASN.1 data specification. For example, an interaction record contains, among other data object, two BIND-objects. These BIND-objects are themselves defined using simpler sub objects. Normally, a BIND-object that is describing a protein sequence will store a simple link to a sequence database, such as GenBank. If, however, the sequence is not present in the public database, it can be fully represented using an NCBI-Bioseq object. The NCBI-Bioseq object is how NCBI stores all of the sequences in GenBank and is a mature data structure.

BIND also inherits the NCBI/EMBL/DDBJ taxonomy model and data, and is designed so that interactions can be both inter and intra organismal. Below is an example of the types of biological data for each BIND record. Explanations of the various objects in the specification are given along with examples. The BIND specification is explained as if it were being used to describe a single record in a database.

The BIND database is generally meant to reference information from other databases rather than storing the information as a copy. This avoids unnecessary duplication of information among databases and helps maintain data integrity (if the information in a referenced record in one database is updated, the other databases that reference the record are all automatically updated). All fields are non-optional unless stated otherwise.

A BIND-object

A BIND-object represents any chemical object - atom, molecule or complex of molecules. See Figures 1 and 2 for a diagrammatic description of the data type. A BIND-object contains:

1. *A short-label field to contain a short name for a molecule.* For example, ATP, IP3, S4 and HSP70 are acceptable short labels for ligands and proteins, respectively. Having a non-optional short label ensures that at least some descriptive data is entered for a molecule. This information is also useful to construct top-level descriptions regarding a particular record. For example, a simple description of an interaction between two proteins can be constructed using the short labels of the two BIND-objects in an interaction record. A graphical view of an interaction would be labeled with the short label field.

2. *A BIND-object-type-id object to contain the type of the molecule and a reference to another database containing a record for that molecule.* In this way, for instance, large DNA records are referenced rather than duplicated. A molecule type may be 'not-specified', 'protein', 'dna', 'rna', 'ligand', or 'molecular complex'. Molecules of unknown type may be stored by specifying the type of molecule as 'not-specified'. This type requires no further data input.

Protein, DNA and RNA all require a BIND-id object. This object can store accession numbers to any other database. It has special fields 'gi' or Geninfo and 'di' or domain identifier for the NCBI Entrez system (Schuler et al., 1996) and a database of domains, respectively. Any other accession number or numbers/strings to reference records in other databases can be stored in a set of NCBI Seq-id's present in the data object. All fields in BIND-id are optional so molecules stored internally in a BIND record that are not present in other databases (and so do not have accession numbers) can be properly saved.

Molecules of type 'ligand' require a BIND-ligand-id object. This object can contain a reference to an internal small molecule database or any other small molecule database via a database name and an integer and/or character based accession number.

BIND-objects of type 'complex' require an integer accession number to a BIND molecular complex record.

3. *A BIND-object-origin data structure.* This data structure contains a choice of origin between 'not-specified', 'org' or organismal, and 'chem' or chemical. BIND-objects of unknown origin would have origin type 'not-specified'. Chemical objects that are derived directly from organisms, such as DNA, would be specified to be origin type 'org' and are required to be associated with an NCBI BioSource object. A BioSource object can contain much descriptive data about an organism and the biological source of a compound. It also contains a reference to a taxonomy database. This information can be entered automatically if a GI is known for a biological sequence molecule, since a BioSource is part of the NCBI Bioseq object which stores biological sequences in Entrez. If a GI is not given, a BioSource can be created.

Molecules derived purely from chemical means are of origin type 'chem' and require a BIND-chemsource object. The BIND-chemsource object contains a set of names for the chemical, usually a common name and any synonyms, a SMILES string (Weininger, 1988), the chemical formula, molecular weight (a RealVal-Units object), and a CAS registry number (http://www.cas.org). A SMILES string is a standard way of representing a molecule's structure using ASCII characters. Many

chemistry computer applications are available to manipulate and use data of this type. Three-dimensional structure of a molecule can be predicted from a SMILES string to a high degree of accuracy using commercial chemistry applications such as Corina (Gasteiger, 1996) and others. A CAS number is a reference number to the information regarding a chemical compound in the Chemical Abstracts Service. This service contains data on at least 22,468,564 chemical compounds. Of all the fields in a BIND-chemsource object, only 'names' is non-optional. This means that for a BIND-object to be declared a ligand of chemical origin, one must only provide a pointer to a small molecule database and one name of the chemical.

4. *An optional BIND-cellstage list to contain a list of cell cycle stages in which this object is found, or expressed, in the given organism.* This information is only relevant for BIND-objects of organismal origin. A BIND-cellstage object is an enumeration of all of the basic cell stages in the cell cycle. It contains an optional text description field that can describe other cell stages that are not present in the enumeration.

5. *An optional NCBI Bioseq object to store a biological sequence if a record for the sequence is not present in any public database.* The Bioseq may also be used to store the experimental form, such as His tagged proteins or mutants, of the biological sequence if it is different from any public database record. This field is only relevant for biological sequences. Bioseqs can be prepared using Sequin (Kans et al., 1998) and can be exchanged with NCBI.

6. *An optional NCBI Biostruc object to store a three dimensional atomic structure of any chemical object, from an atom to a complex of molecules, if the data is not present in any public database.* The Biostruc specification allows a chemical graph to be stored without coordinates. This is most useful for storing small molecule structures or post-translationally modified forms of a biomolecule. Thus, chemical entities within a BIND object can be described in precise detail.

The presence of these powerful and mature data structures in this part of the specification demonstrates that BIND is not completely reliant on other databases. Most of the information present in any public sequence or 3D molecular structure database can be stored using the BIND specification if necessary.

7. *An optional free flow text description of the BIND-object.* This field contains, for example, a full name for a molecule such as Adenosine Triphosphate (ATP).

BIND-Interaction

The BIND-Interaction object is the fundamental component for storing data in this specification. It defines and describes the interaction between any two molecules, or even atoms. The majority of the information that can be stored is, however, used to describe interactions between proteins, DNA and RNA. Interactions between molecules rather than between molecules and atoms are exemplified from this point on. See Figure 3 for a diagrammatic representation of the data type.

A BIND- interaction contains a NCBI Date object, a sequence of updates for an audit trail, an Interaction Identifier (IID) accession number, two interacting molecules (BIND-object), a description of the interaction, a series of publications and a private flag. BIND IID number space is to be

controlled using a unique key server. Molecule A binds to molecule B and both are stored using BIND-objects (described above).

The BIND-descr object stores most of the information in an interaction object. It contains text description of the interaction, information on cellular place of interaction, experimental conditions used to observe the interaction, conserved sequence comment of molecules A and/or B if they are biological sequences, location of binding sites on molecule A and B, chemical actions mediated by the interaction and chemical states of the molecules A and B.

A BIND-pub-set is included to store empirical evidence references, usually publications, that 'support', 'dispute' or have 'no opinion' regarding the actual interaction. The dispute flag allows the database to track experimental trends and offer a machine-readable way to find discrepancies or differences of opinion.

Finally, the private flag which defaults to FALSE is included in the BIND-interaction record. The flag indicates whether or not to export this record during a data exchange procedure. In a public database, a private record is not available to the public. This may be because the record has not been completed or information in the record has not been verified. In a private database, the private flag means that the record can be viewed internally, but not exported. In this situation, a private record might contain proprietary information. BIND may contain a mix of these and public records imported from a public database.

Interaction description - BIND-descr

All of the objects directly linked in this structure are optional to allow any level of richness of data to be stored. BIND-descr contains:

1. *A simple text description of the interaction.* This free flow text is meant to be a short description of the interaction such as, "transcription factor X binds to a region of human DNA in section x of chromosome 11".

2. *A sequence of BIND-place objects.* See Figure 4 for a diagram of this data type. A BIND-place object stores information about the location of the interaction with respect to the cell. The place of an interaction is meant to be the location where molecule A and B come together in a biologically meaningful way. This object contains a BIND-gen-place-set object for storing general place data, an optional BIND-spec-place-set object for storing specific place data, an optional BIND-pub-set for storing publications referring to the localization of an interaction, and an optional text description field. A BIND-gen-place-set contains a start and an optional end place for the interactions, specified by an enumerated list of general places in the cell. Storing a start and an end place for an interaction takes into account the possibility of an interaction translocating across membranes and ending up in different sub-cellular compartments. The general enumeration of cell places allows a computer to understand the location of the interaction. Only basic cell places are present in the list. This is important for data visualization programs that need to be able to draw molecules in the correct places on a diagram of a cell. A human readable description of cellular place can be stored in the BIND-spec-place-set. This object contains a text description of a start and an optional end place for an interaction. More specific

data regarding the location of interaction, such as in what part of a membrane, apical or basal, an interaction occurs can be stored in the BIND-spec-place-set object.

Multiple BIND-place objects are present to allow storage of an interaction that may be present only at certain separate places within and around the cell. More than one BIND-place object can also be used to describe an interaction occurring between two molecules over multiple sub-cellular compartments, as might be the case for transmembrane receptor proteins with large extra and intra-cellular domains.

3. *A BIND-condition set to store a list of experimental conditions used to observe the interaction.* See Figure 5 for a diagrammatic view of the BIND-condition data type. Experimental conditions information stored should be sufficient to allow recreation of the original experiment. An experimental condition is described using a BIND-condition object. This object contains an Internal-conditions-id (ICID) number which can be used to reference a particular experimental condition in the BIND-condition-set. A general experimental condition is an enumeration of three general conditions, *in-vitro*, *in-vivo* and other. A BIND-experimental-system object is present and is an enumeration of most popular experimental techniques, with 34 techniques listed in the specification. This field has been simply declared as an INTEGER enumeration type so that it can be easily extended with new experimental systems as they become available. Declaring a type as INTEGER in ASN.1 instead of enumeration prevents generated code from checking the name of the enumerated value against the specification. This means that items may be added to the list at a later date without disrupting tools that are based on previous specifications. A BIND-condition object also contains a free human readable text description. This field could be used to describe a system further or could be used to name a system if 'other' has been specified as the BIND-experimental-system object. A BIND-pub-set is also provided in order to store publications related to the experimental systems described in the BIND-condition object.

4. *A BIND-cons-seq-set to store information about evolutionarily conserved sequence if either molecule A or B is a biological sequence.* See Figure 6 for the diagram of this data object. This information is simply meant to be a comment on the possible importance of certain sequence elements that have been noticed to be conserved via phylogenetic or other evolutionary analysis. It is possible that information about conserved sequence is known for molecules in an interaction that is not very well characterized. This data might be useful to investigators interested in further studying the interaction. A BIND-cons-seq-set contains conserved sequence information about molecule A and B in a BIND-conserved-seq object. Semantically, a BIND-conserved-seq object may only be instantiated with data if the molecule that it refers to is a biological sequence. A BIND-conserved-seq object contains an NCBI Seq-loc object. A Seq-loc can contain a location or a set of locations for any linearly numbered biological sequence. A free text description is also included in a BIND-conserved-seq. It is suggested that the method of determining the conserved sequence, for example a phylogenetic tree program such as PHYLIP (http://evolution.genetics.washington.edu/phylip.html) or an alignment program such as PSI-BLAST (Altschul et al., 1997) or CLUSTAL (Higgins et al., 1996) be stored in

the 'descr' field. A BIND-pub-set object is provided to store publications pertaining to a conserved sequence comment.

5. *A BIND-loc to store binding site information.* Figure 7 contains a diagrammatic view of this data type. The BIND-loc can store 3D atomic level detail of an interaction site using an NCBI Biostruc. A BIND-loc-gen object is present to store binding sites in an interaction at the sequence element level of detail. Therefore, only interactions involving biological sequences can hold general binding site information. The BIND-loc object also includes a BIND-pub-set for storing publications related to binding site. All top level fields are optional allowing detailed, general and/or source information to be represented. Expanding further, the BIND-loc-gen object contains a list of binding sites on molecule A and a list of binding sites on molecule B. This information is contained in a BIND-loc-site-set object which contains a sequence of binding sites defined in BIND-loc-site objects. Each BIND-loc-site element contains an NCBI Seq-loc element and an internal reference integer ID called a BIND-Seq-loc-id. Since each binding site is numbered in a BIND-loc-site-set, it can be referenced by other objects.

A BIND-loc-gen object also contains an optional BIND-loc-pair object which specifies which binding sites on A bind to which binding sites on B. The binding sites are referenced from the BIND-loc-site-set objects so in order to use a BIND-loc-pair object, binding sites on molecule A and B must already be defined. This simple binary mapping allows most experimental binding information, such as that generated from footprinting analysis, to be stored.

6. *A set of BIND-actions to describe the chemical action(s) mediated by this interaction.* Figure 8 shows a diagram of this data type and related objects. A set of actions is required because there are many examples of interactions having multiple chemical actions. For instance, a kinase may phosphorylate a protein more than once in separate chemical actions or a restriction enzyme may cleave a molecule of DNA in more than one place. A BIND-action-set contains a set of elaborate BIND-action objects. Each BIND-action object in a set is numbered with an Internal-action-id (IAID) integer so that it can be referenced by other data types.

A BIND-action object contains an IAID number, an optional text description field for free flow text description of the chemical action and an optional BIND-pub-set for storing publications pertaining to this chemical action. A boolean flag is included to specify the direction of the chemical action. If a-on-b is set to true, then molecule A acts on molecule B, and vice versa. This value defaults to true. The type of action is defined in the BIND-action-type object. The BIND-action-type object is a choice element that stores the type of chemical action and an associated data object. The possible choices of actions are 'not-specified' for an unknown chemical action type, 'add' for adding a chemical object, 'remove' for removing a chemical object, 'cut-seq' for a cut in a biological sequence, 'change-conformation' for a change in conformation, 'change-configuration' for a change in configuration, e.g. by an epimerase or isomerase, 'change-other' for another type of change, such as a metal ion exchange, and 'other' for any other chemical action. Types 'add', 'remove' and 'cut-seq' are associated with a BIND-action-object to store related data.

A BIND-action-object is a choice element that can store nothing, with a choice of NULL, a BIND-object, or a site on a sequence using a Seq-loc. The 'object' choice of the BIND-action-object is only relevant for the 'add' and 'remove' choices of the BIND-action-type. The BIND-object is meant to store a description of the chemical compound that is added or removed. An example would be a phosphate group that could be added by a kinase enzyme or removed by a phosphorylase enzyme. The 'location' choice of the BIND-action-object is only relevant for the 'cut-seq' choice of the BIND-action-type. The Seq-loc is meant to store the position(s) where a biological sequence is cut. An example would be the locations after which a restriction enzyme cuts DNA or the sites after which a protease cleaves in a protein. The choice of 'none' can be used for either 'add', 'remove' or 'cut-seq' if information that would otherwise be stored is not known.

The BIND-action object also includes an optional result field to store the resulting molecule(s) from a chemical action as a sequence of BIND-objects. For instance, if a molecule of DNA was methylated, the description of the methylated DNA could be stored in a BIND-object. If a protein molecule was cut at various locations, all resulting protein molecule fragments could be described with the BIND-object sequence. With a sequence of interacting proteins where A binds to B, B binds to C, etc., the result field storing the full chemical form of B in the A-B interaction, for example, could be used directly in the B-C interaction record. This allows the exact description of sequential chemical modifications on a biological sequence that would otherwise not be possible given the standard sequence representation alone.

A Biostruc-feature-set that can contain residue or atomic level of detail differences in a molecule created by this chemical action is also present as a BIND-action object. The molecule that is different in this case is based on the direction of the chemical action. If the direction is molecule A to B, any information stored in the diff field would pertain to molecule B, not A. This field allows even small changes to molecules to be represented, as in the example of a chemical action reducing a double bond by adding two hydrogen atoms across it. The addition of the two hydrogen atoms could be recorded as differences on an atomic structure. This information requires the presence of atomic level detail data for the molecule being changed. The diff field can also represent changes made to the substrate of the chemical action. In an example of a phosphate added to a protein on a specific tyrosine residue by a phosphokinase enzyme, the diff field would simply be the position in the protein sequence of the tyrosine that was being changed.

An optional BIND-signal object is included in the BIND-action object to store directional information related to chemical signal as it is found in cell signaling pathways. This data is really a more general notion of kinetics describing signal transduction. The signal could, for example, be the activation of proteins in a signaling cascade via phosphorylation such as in a MAP kinase pathway. BIND-signal object contains an enumerated type describing the signal modification from a top-level viewpoint. Possible values are 'none', 'amplify', 'repress', 'auto-amplify', 'auto-repress', and 'other'. The direction of the signal is stored in the a-to-b boolean flag, which defaults to true. If a-to-b is true, the direction of signal is from molecule A to molecule B and vice versa. An optional RealVal-Units field

can store the factor of signal amplification or repression if they occur. Signal amplification in the cell is really just the recruitment of molecules one step further down in the pathway by the molecule at the current step. So, if molecule A activates molecule B by removing a phosphate in a signaling pathway and there is amplification at this step, in the cell, molecule A activates many molecules of B causing a strengthening of the chemical signal by a measurable factor that may be stored. An optional free text description is available in the BIND-signal object as well. This field should contain some description of the signal action if 'other' is specified in the 'action' field.

Kinetic and thermodynamic data may also be optionally stored in the BIND-action object using the BIND-kinetics object. The BIND-kinetics object offers specified real value and text description fields for common kinetics (e.g. Michaelis-Menten) and thermodynamic values as well as providing a sequence of BIND-kinetics-other objects to store any other text or real number values that may be pertinent. A BIND-pub-set object is also present to store publications that relate to any of the information stored. All objects in the BIND-kinetics object are optional to allow any combination of values to be stored.

Also in the BIND-action object, a link to a sequence of experimental conditions used to observe this chemical action is optionally provided using a sequence of BIND-condition-dependency objects. The BIND-condition-dependency objects reference previously defined experimental conditions by Interaction-id and Internal-conditions-id number. In this way, any experimental condition in a database using this specification may be uniquely referenced.

7. *A BIND-state-descr object for storing information on chemical state of molecule A or B.* See Figure 9 for a diagrammatic view of this data type. The BIND-state-descr object stores a list of possible chemical states for molecules A and B in BIND-state-set objects as well as references to defined chemical states of A and B that are required for the interaction to take place, in BIND-required-state objects. More than one possible state can be saved because certain molecules can assume multiple states. One example is a protein enzyme which may be multiply phosphorylated to bring about different enzymatic activity levels, depending on the phosphorylation level. All fields in the BIND-state-descr object are optional allowing any combination of data objects to be stored. A BIND-state-set contains a sequence of BIND-state objects each numbered by an Internal-state-id (ISID) integer. Each BIND-state object contains an optional enumerated list describing the general activity of the molecule, an optional sequence of BIND-state-cause objects, an optional free text description, and an optional BIND-pub-set for storing publications related to this chemical state description. The 'activity-level' list is a simple description and is purely subjective, but is still useful for discriminating various states of different activity, especially by a data visualization program which could colour molecules based on this information.

The BIND-state-cause object can be used to uniquely reference previously defined chemical actions from this or other interactions that bring about this state. It contains an IID and an IAID. This functionality is very important in the specification because it allows full chemistry to be described when chemical actions and chemical states are taken together. Full chemistry means that all substrates,

enzymes, products, bio-processed compounds etc. may be represented in full atomic level detail for all steps in a pathway. A certain chemical action can have a result (in the 'result' field of a BIND-action object) and a certain chemical state can reference the action that occurred to create it. In this way bi-directional linked lists can form networks that represent true chemical networks in a cell.

A Molecular Complex - BIND-Molecular-Complex

The BIND-Molecular-Complex object is the second of three top-level biological objects in the BIND specification. It is meant to store a collection of more than two interactions that form a complex, i.e. three or more BIND-objects that can operate as a unit. In this way, it is useful to store knowledge of molecular complexes and as a shorthand for use when defining interactions and pathways (see BIND-pathway). Figure 10 provides a box diagram view of this data type.

A BIND-Molecular-Complex object contains similar administrative information fields as a BIND-Interaction object. A Molecular-Complex-id (MCID) integer accession number is stored to uniquely identify molecular complexes. A BIND-pub-set is present to store publications that concern this molecular complex and a private flag is provided to mark this record as private using the same rules as the private flag of the BIND-interaction record.

Six other fields in the molecular complex store data directly relating to the complex. A 'descr' field optionally provides space for a human readable free text description of the molecular complex. The 'sub-num' field contains a BIND-mol-sub-num object that stores the number of sub-units (BIND-objects) in the molecular complex. The sub-unit number includes either an exact integer using the 'num' field or a fuzzy integer in the 'num-fuzz' field. The fuzzy number is stored using an NCBI Int-fuzz object which can store a number in a range, plus or minus a fixed or percentage amount, or store a set of alternatives for the number. Using a fuzzy number, complexes can be stored even when the exact number of sub-units is not known. Examples of such complexes are actin filaments or other parts of the cytoskeleton and virus coat proteins, both of which typically form using repeated units of a certain protein.

The BIND-Molecular-Complex object also includes a 'sub-units' field to store the actual sub-units of the complex as a sequence of BIND-mol-object data types. The BIND-mol-object is simply a wrapper for a BIND-object that allows the BIND-object to be numbered using a BIND-mol-object-id integer (BMOID). Numbering the sub-unit BIND-objects allows the BIND-mol-object-pair to reference them for topology, as discussed below.

A primary component of the BIND-Molecular-Complex object is a list of Interaction-ids, which references previously defined interactions in a database. This means that most of the data for function, state, location, etc. for a molecular complex is actually stored in BIND-Interaction objects. This avoids some duplication of information. A boolean flag marks the interaction list as being ordered or not. This should be true if the temporal order of interactions that form the complex is known and the IID list is ordered in that way. Ordering of sub-unit binding for some well studied biological complexes, such as the ribosome, is known.

An optional sequence of BIND-mol-object-pair objects is present in the BIND-Molecular-Complex object and is meant to store a two-dimensional topology of the molecular complex. A BIND-mol-object-pair object simply records a connected pair of BIND-mol-objects in the molecular complex by making a reference to two BMOID numbers of the sub-units that are connected. Together the BIND-mol-objects, as nodes, and the BIND-mol-object-pair objects, as edges can describe the computer science concept of a graph. The topology information can allow a data visualization program to draw a representation of the actual shape of the complex.

Because most of the data for complexes is referenced BIND-interaction records, a certain amount of automatic data entry can be used. A list of sub-units and the number of sub-units can be automatically entered by fetching the data from the given list of interaction records.

It can also be noted that a molecular complex can be defined if the pairwise interactions of which it is composed are not completely known. This can be done by creating a set of interaction objects with molecule A as a sub-unit of the complex and molecule B as 'not-specified'. This is useful since many preliminary studies of a molecular complex observe only that certain molecules interact, e.g. from gel data, but not how they interact.

A Pathway - BIND-Pathway

The final top-level biological object in the BIND specification is the BIND-pathway data type. It describes a collection of more than two interactions that form a pathway, i.e. three or more BIND-objects that are generally free from each other, but can form a network of interactions. Common examples include metabolic pathways and cell signaling pathways. See Figure 11 for the box diagram for this data type.

A BIND-Pathway object contains similar administrative information fields as a BIND-Interaction and a BIND-Molecular-Complex. Two other fields in the BIND-pathway object store information describing the pathway. A sequence of Interaction-ids that reference previously defined interactions that make up this pathway is stored. Extra descriptive information regarding the pathway is stored using a BIND-path-descr object. This object can optionally store free text describing the pathway and an optional sequence of BIND-cellstage objects that represent the phases of the cell cycle in which this pathway is in effect. Parts of the pathway may be constitutively present in the cell, while other parts that complete the pathway and allow activation may only be expressed at certain times during the cell cycle.

Other BIND ASN.1 objects

Publication Set

A BIND-pub-set is used to hold all publications in BIND. It contains a list of BIND-pub-objects and a dispute flag. A BIND-pub-object contains an optional free text description of the publication, an enumerated opinion of the publication field and a NCBI Pub object. The description field may hold any text data pertaining to the publication referenced by this object. The opinion field may hold the values: 'none', 'support' and 'dispute'. It is meant to convey the general opinion of the referenced publication in regard to the information in the ASN.1 object that contains the BIND-pub-set.

The NCBI Pub object is used to store most of the data in PubMed and can represent almost any publication. It should be used to store a reference to PubMed whenever possible using either a Medline Unique Identifier (MUID) or a PubMed unique identifier (PMID).

Record Update

If a record is updated in BIND, a description of the update should be added to a BIND-update-object. This object contains a NCBI Date object and a text description field. The description field may contain any information that a database implementation decides to store, but it should be complete and stored in a standard and automatic way within each implementation so that it can be easily parsed. Any information may be stored up to and including the entire previous record in ASN.1 value notation. This data is not meant to be human entered but rather maintained as a machine generated audit trail of any changes made.

Data exchange and data cross-referencing

Data exchange systems and database management data structures have been included in the specification as powerful tools to make implementations more robust. BIND-Submit is the top-level object for data exchange while the cross referencing system involves many separate top-level data objects.

Data exchange - BIND-Submit

The BIND-Submit object can be used to exchange any number of the top-level data types in the BIND specification, BIND-Interaction, BIND-Molecular-Complex, and/or BIND-Pathway objects. BIND-Submit stores an NCBI Date object, an optional BIND-Database-Site, a BIND-Submitter object, an optional BIND-Submit-id integer for identifying this submission, and fields for optionally storing BIND-Interaction-set, BIND-Complex-set, and BIND-Pathway-set objects.

A BIND-Database-site is a description of a database site. This object could be used if data was being submitted to BIND from any other database. It contains free text description of the database site, usually the database name. Also present is a text field for database country of origin and an optional field used to store the World Wide Web Universal Resource Locator (WWW URL) of the homepage of the database on the Internet. An optional NCBI Pub object can store a Medline reference for this database.

A BIND-Submitter object contains information about a submitter to a BIND database. BIND-Submitter stores a BIND-Contact-info object which contains information about a person. A "hold until published" boolean flag is present which defaults to false to allow data submission prior to publication. Also present is an optional enumeration of possible submission types, either 'new', 'update', 'revision', or 'other'. An update is a change by an author while a revision is a non-author update. A free text field, 'tool', stores the name and version of the tool used to submit the record.

Personal contact information may be kept separate from BIND records to keep the submitter and ownership information anonymous and protected from improper use.

Actual records are stored in the BIND-Submit object in data set data types. The BIND-Interaction-set, BIND-Complex-set and BIND-Pathway-set are all present in the BIND-Submit object

and are analogous in that they optionally store the date on which the set was collected, optionally the database from which the record set originates using a BIND-Database-site, and the respective sequence of records.

Cross-Referencing the Data

Since the BIND specification describes biological data from interactions to pathways and networks of pathways, the information space represented resembles a largely undirected graph with molecules as vertices and their interactions as edges. Cross-referencing information allows the graph to be easily traversed using simple indexed lookup techniques. If cross-referencing were not used in a system such as this, all records would have to be examined at each traversal of the data space. Instead of creating traditional large, unwieldy indexes and tables to speed the traversal process, ASN.1 objects are directly specified to store cross-reference information. This represents an object oriented database index system. Each BIND database accession number as well as NCBI GI, MUID and PMID and SLRI DI accession numbers has its own associated cross-reference object. This information may be easily exported and used by other databases to link their sequence or structure data back to BIND.

When updating cross-reference information, only one level of the graph is traversed, so as not to make the index overly complicated. Any time one of the three top level objects is created that contains a cross-referenced accession number, the BIND-Cross-Ref object lists are updated. In this way, any search using a cross-referenced accession number instantly retrieves all of the interaction, complex and pathway records that contain it.

The interaction cross-reference data is stored in a BIND-Iid-Cross-Ref object. This data type contains the IID of the interaction being cross-referenced in this object. The 'iids', 'pids' and 'mcids' fields contain a list of IIDs, PIDs and MCIDs, respectively of interactions, pathways and complexes that contain this interaction. A BIND-Submitter object is included to privately store submitter information for every interaction.

Molecular complex cross-reference information is stored in a BIND-Mcid-Cross-Ref object which is completely analogous to the BIND-Iid-Cross-Ref object.

Pathway cross-reference data is contained in a BIND-Pid-Cross-Ref object. This object only keeps a list of submitters for each pathway record. Since no other objects can reference a pathway record, the BIND-Pid-Cross-Ref object does not contain references to other records.

The GI/DI cross-reference information is stored in a BIND-Cross-Ref object. This object links a biological sequence to a list of interactions, molecular complexes and pathways that contain it.

PMID/MUID cross-reference data is maintained in a BIND-Pub-Cross-Ref object. This cross-reference scheme is analogous to that of GI/DI accession numbers.

The full cross-reference system allows quick and easy searching of the database by any of the five indexed accession numbers.

Exported data types

Typical ASN.1 data specifications make certain data types available for use by other ASN.1 specifications by exporting them. BIND currently exports the top-level data types BIND-Submit,

BIND-Interaction, BIND-Interaction-set, BIND-Pathway, BIND-Pathway-set, BIND-Molecular-Complex and BIND-Complex-set.

Flat-file Record Format

Many current biological data specifications are available in a flat-file format for use with simple flat-file databases. Examples include the GenBank flat-file format and the FASTA format for biological sequence representation. BIND can be made available in a flat-file database record format that mirrors the BIND ASN.1 specification. Therefore, BIND may include ASN.1⇔flat-file conversion software tools.

Data Entry

BIND may rely on the following different sources for data entry.

1. Manual data entry:

Data is entered manually via web based forms handled by CGI scripts on a World Wide Web Server. This allows entry of data from individual computers on a users own time, from anywhere in the world. BIND indexers review and validate public entries as they arrive. Researchers can enter their data after they have finished an experiment.

(i) Curated Data entry:

Data that is already present in the literature will be entered into BIND.

(ii) User data entry:

When a paper about protein or DNA sequence or protein structure is about to be published, an author generally obtains an accession number to a database, such as GenBank or PDB. An author of a paper containing information about biomolecular interactions, complexes, or pathway information, will obtain a similar accession number from BIND. A BIND indexer will validate the incoming data and issue an accession number. This follows the GenBank model.

2. Automated data entry:

Data gathering agents will gather data from various sources on the Internet. Possible examples include:

A. NCBI's MMDB structure database that contains many protein multimers, with accompanying detailed atomic interaction information. Web site: www.ncbi.nlm.nih.gov/Structure/

B. DIP (Database of Interacting Proteins) contains many protein-protein interactions. Web site: http://ampere.mbi.ucla.edu:8801/

C. FlyNets – the drosophila interactome database – contains protein-protein, protein-DNA and protein-RNA interactions. Web site: http://gifts.univ-mrs.fr/GIFTS_home_page.html

D. Ligand DB – compound database from Japan. Web site: www.genome.ad.jp/dbget/ligand.html

E.Klotho–another compound database. Web site: www.ibc.wustl.edu/moirai/klotho/compound_list.html

3. Data entry direct from experimental systems:

A separate instance of BIND, a BIND satellite database, can be used as a local repository to store experimental data as it is gathered, but before it is analyzed. Any data that is then used in a publication can then be transferred easily to the public database. A BIND satellite can download the current public database and merge it with local data.

Examples of experiments that can be used to locate interactions include:

A. Immuno-precipitation

B. Affinity chromatography

C. Yeast two hybrid

D. DNA footprinting

E. Reconstitution experiments (using various detection tools such as FRET, hydroxyl radical footprinting, isotope exchange combined with mass spectroscopy, and fluorescence anisotropy)

Accessing BIND Data

BIND can be accessed via a user-friendly Web interface on the Internet and anyone using a current web browser can access BIND data. BIND records may be searchable by Interaction ID (iid), Molecular Complex ID (mcid), Pathway ID (pid), NCBI gi, and PubMed or Medline ID. The data can be text indexed and searchable using keywords. There is a BLAST interface to BIND.

Visualization of BIND data:

Web based Java applets that will dynamically represent pathways and molecular complexes have been designed. These form the preferred front end of the BIND system. For example, when a pathway is graphically represented, the image is mouse clickable so that information about the record and other records in the database will be easily accessible.

Implementation

This section gives an overview of the BIND database. The implementation allows data entry and data retrieval supporting the full BIND 1.0 ASN.1 specification. Programmed fully using the C programming language for maximum speed and compatibility, a BIND application programming interface (API) has been written to allow applications to easily use data in the BIND database. The API makes use of two C libraries, the NCBI Toolkit (ftp://ncbi.nlm.nih.gov/toolbox) for ASN.1 handling and more and the CodeBase (http://www.sequiter.com/) database library for a database implementation. Using this API, web-based applications have been developed for data entry, retrieval and management. All data is entered and retrieved using web-based forms generated by CGI programs written in C. Interaction data is entered using this web-based user interface.

The BIND database uses the Seqhound database system as a resource. Seqhound is a mirror of GenBank, the NCBI taxonomy database and the PDB (Bernstein et al., 1978) data in NCBI MMDB form (Hogue et al., 1996). Seqhound derived data allows BIND to quickly and easily use sequence, taxonomy and 3D molecular structure information for validation and for information retrieval.

Data visualization and data mining systems have been designed for the database implementation. The spider will traverse BIND searching for new signaling pathways. It will traverse all pathway cross talk links looking for signaling routes that are not present in BIND. The results of

this search will be potentially unknown cellular signaling pathways. The proteins of these new pathways can also be examined to see if they contain known binding domains, such as SH2 and SH3 domains, which will increase the likelihood of pathway cross talk. The short list of newly found potential pathways can then be experimentally evaluated.

With the current information garnered by genomic sequencing projects, homologous cell signaling pathways can be found in other organisms by knowing all of the gene products in a pathway in a related organism. Even between non-related organisms, certain 'housekeeping' pathways should not be expected to differ much.

BIND can be used to find networks of biological signaling pathways whose topologies can support signal properties that simple pathways can not. It has been shown that certain kinds of signaling networks have properties that cannot be seen with simple signal pathways. Storing of information, large-scale signal attenuation and signal control are some of these properties. It has been supposed that memory can have a basis in the long term storing of information in certain signaling pathways. (Bjalla, US and Iyengar R, Emergent Properties in Signaling Networks, Science 283(5400):381-7, Jan 15, 1999)

BIND can also be used to identify a biomolecular interaction that is similar to a reference biomolecular interaction stored in BIND. The user interface allows the user to initiate a similarity search for each molecule in the test biomolecular interaction. The results can be screened by selected taxonomy. A putative biomolecular interaction is then assembled to create a test record. The BIND database is then examined to match the test record with the records therein to produce a matching record containing a reference biomolecular interaction that matches the test biomolecular interaction.

Examples of a BIND specification are attached as Appendix A and Appendix B.

Having illustrated and described the principles of the invention in a preferred embodiment, it should be appreciated to those skilled in the art that the invention can be modified in arrangement and detail without departure from such principles. We claim all modifications coming within the scope of the following claims.

All publications, patents and patent applications referred to herein are incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

APPENDIX A

```
--$Revision: 0.5 $
-- ********************************************************************************
--
-- BIND (Biomolecular Interaction Network Database) Interaction Record
-- by Gary Bader. Oct. 21, 1998
-- Hogue Lab - University of Toronto Biochemistry Department
-- Samuel Lunenfeld Research Institute, Mount Sinai Hospital
--
-- ********************************************************************************

BIND-Interaction DEFINITIONS ::=
BEGIN

EXPORTS BIND-Interaction, BIND-interaction-set,
        BIND-Pathway, BIND-pathway-set,
        BIND-Molecular-Complex, BIND-complex-set;

IMPORTS Date FROM NCBI-General
        Bioseq FROM NCBI-Sequence
        Submit-block FROM NCBI-Sumit
        Pub FROM NCBI-Pub
        Org-ref FROM NCBI-General
        Seq-loc, Seq-id FROM NCBI-Seqloc
        Biostruc FROM MMDB
        Biostruc-graph FROM MMDB-Chemical-graph
        Chem-graph-pntrs FROM MMDB-Features;

-- ***************
-- * Interaction *
-- ***************

-- **********************
-- A set of interactions
-- **********************


BIND-interaction-set ::= SEQUENCE OF BIND-Interaction


-- ********************************************************************************
-- A BIND-Interaction record can store all of the details of the interaction
-- between any two molecules (or atoms).
--
-- Field description for BIND-Interaction
-- ****************************************
-- date = date of record entry
-- updates = a list of updates for the record
-- iid = interaction accession number
-- pids = list of pathways that this interaction is involved in
-- mcids = list of molecular complexes that this interaction is involved in
-- a = molecule 'a' interacts with...
-- b = molecule 'b'
-- descr = description of interaction
-- source = empirical evidence references
-- sub = contact information of the submitter and general information about
--       this submission
-- priv = TRUE if this interaction is private
-- ********************************************************************************
```

```
BIND-Interaction ::= SEQUENCE {
        date Date,
        updates BIND-update-set OPTIONAL,
        iid Interaction-id,
        pids SEQUENCE OF Pathway-id OPTIONAL,
        mcids SEQUENCE OF Molecular-Complex-id OPTIONAL,
        a BIND-object,
        b BIND-object,
        descr BIND-descr,
        source BIND-pub-set,
        sub Submit-block,
        priv BOOLEAN DEFAULT FALSE
        }

Interaction-id ::= INTEGER

Pathway-id ::= INTEGER

Molecular-Complex-id ::= INTEGER


-- *****************
-- * Record Update *
-- *****************

-- ******************************
-- A set of updates for a record
-- ******************************


BIND-update-set ::= SEQUENCE OF BIND-update-object


-- ***********************************
-- An update for a record
--
-- Field description for BIND-update
-- ***********************************
-- date = date of this update
-- descr = text description of update
-- ***********************************


BIND-update-object ::= SEQUENCE {
        date Date,
        descr VisibleString
        }


-- ************************
-- * Biomolecular Object *
-- ************************

-- ***************************************************************************
-- Any chemical object
--
-- Field description for BIND-object
-- *********************************
-- id = a choice of possible pointers (usually to accession numbers
--      of other databases) for different types of molecules that may
```

```
--          interact.
--          Choice of: none, protein, dna, rna, ligand (any other type of
--                      chemical compound), interaction, molecular complex
-- origin = material source, org-ref for biological material, text
--             description for chemical compound of non biological origin
-- seq = space for sequence, if it is not in a public database
-- struc = space for complete structure, if not in public database
-- descr = optional text description of this object
-- *********************************************************************


BIND-object ::= SEQUENCE {
        id CHOICE {
                none NULL,
                interaction Interaction-id,
                complex Molecular-Complex-id,
                protein BIND-id,
                dna BIND-id,
                rna BIND-id,
                ligand BIND-ligand-id
                },
        origin CHOICE {
                org Org-ref,
                chem VisibleString
                },
        seq Bioseq OPTIONAL,
        struc Biostruc OPTIONAL,
        descr VisibleString OPTIONAL
        }


-- ***************
-- * Identifiers *
-- ***************

-- ************************************************************************
-- General sequence or domain identifier
--
-- Field description for BIND-id
-- *****************************
-- gi = NCBI main accession number
-- di = domain accession number (from the domain split database)
-- other = open field for other possible NCBI defined pointers
--
-- NOTE: there is a field for gi in 'other Seq-id', but it should not be used
-- in this specification
-- ************************************************************************


BIND-id ::= SEQUENCE {
        gi Geninfo-id,
        di Domain-id OPTIONAL,
        other Seq-id OPTIONAL
        }

Geninfo-id ::= INTEGER

Domain-id ::= INTEGER

-- ************************************************************************
```

```
-- Pointers to various ligand databases (needs to be expanded e.g. CAS reg.#)
--
-- Field description for BIND-ligand
-- ********************************
-- internal = an accession number describing an internally kept structure of a
--            chemical compound (composite database of LIGAND DB and Klotho DB)
-- other-db = generic pointer to any other database (e.g. Japanese ligand db)
--          Contains the name of the database, an integer pointer and a string
--          pointer.
-- ************************************************************************

BIND-ligand-id ::= CHOICE {
          internal Internal-ligand-id,
          other-db BIND-other-db
          }

Internal-ligand-id ::= INTEGER

BIND-other-db ::= SEQUENCE {
          dbname VisibleString,
          intp INTEGER OPTIONAL,
          strp VisibleString OPTIONAL
          }


-- ****************
-- * Publications *
-- ****************

-- ************************************************************************
-- This holds a publication set
--
-- Field description for BIND-pub
-- ******************************
-- disputed = TRUE if the interaction is disputed in the pub-set
-- pubs = a sequence of pub-objects
-- ************************************************************************

BIND-pub-set ::= SEQUENCE {
          disputed BOOLEAN DEFAULT FALSE,
          pubs SEQUENCE OF BIND-pub-object
          }


-- ************************************************************
-- A publication
--
-- Field description for BIND-pub-object
-- *************************************
-- descr = optional text description of this object
-- opinion = does this publication support or dispute the data
-- pub = publication reference
-- ************************************************************

BIND-pub-object ::= SEQUENCE {
          descr VisibleString OPTIONAL,
          opinion ENUMERATED {
```

```
                none (0),
                support (1),
                dispute (2)
                },
        pub Pub
        }


-- ****************************
-- * Interaction Description *
-- ****************************

-- ********************************************************************************
-- Full description for an interaction
--
-- Field description for BIND-descr
-- ********************************
-- simple-descr = text description
-- loc = description of location of interaction
-- cond = binding conditions/experimental conditions
-- cons = conserved sequence element comment
-- interaction = graph of the actual atoms or sequence elements
--                involved in the interaction
-- action = list of chemical actions that can occur in this interaction
-- state = list of active states of 'a' and 'b' as well as required
--          activity states for interaction to occur
-- ********************************************************************************


BIND-descr ::= SEQUENCE {
        simple-descr VisibleString OPTIONAL,
        loc BIND-loc OPTIONAL,
        cond BIND-condition-set OPTIONAL,
        cons BIND-cons-seq-set OPTIONAL,
        interaction Biostruc-graph OPTIONAL,
        action BIND-action-set OPTIONAL,
        state BIND-state-descr OPTIONAL
        }


-- *************************
-- * Interaction Location *
-- *************************

-- ********************************************************************************
-- Top level interaction location descriptor
--
-- Field description for BIND-loc
-- **********************************
-- gen-loc = general cellular location where this interaction takes place
--        (computer readable)
-- spec-loc = a fully specific location of the interaction
--        (human readable)
-- source = empirical evidence references
-- descr = text description
-- ********************************************************************************

BIND-loc ::= SEQUENCE {
        gen-loc BIND-gen-loc-set,
        spec-loc BIND-spec-loc-set OPTIONAL,
```

```
        source BIND-pub-set OPTIONAL,
        descr VisibleString OPTIONAL
        }

-- ************************************************************************
-- General start and end locations for an interaction
--
-- Field description for BIND-gen-loc-set
-- ************************************
-- start = general location where this interaction takes place
-- end = general location where this interaction ends
-- descr = text description of this object
-- ************************************************************************


BIND-gen-loc-set ::= SEQUENCE {
        start BIND-gen-loc,
        end BIND-gen-loc OPTIONAL,
        descr VisibleString OPTIONAL
        }

-- ************************************************************************
-- General cellular location where this interaction takes place
--
-- Field description for BIND-gen-loc
-- **********************************
-- An enumeration of general cell locations
--        extracellular = extracellular
--        cytoplasm = in cytoplasm
--        organelle = in an organelle
--        nucleus = in nucleus
--        membr-cell-cyt = on the cytoplasmic side of the cell membrane
--        membr-cell-in = in the cell membrane
--        membr-cell-ext = on the surface of the cell membrane
--        membr-outer-peri = on the periplasmic side of the outer membrane
--        membr-outer-in = in the outer membrane
--        membr-outer-ext = on the surface of the outer membrane
--        cellwall-cell = on the inside surface of cell wall
--        cellwall-in = in the cell wall
--        cellwall-ext = on the outside surface of the cell wall
--        other = other location - see text description in BIND-gen-loc-set
-- ************************************************************************

BIND-gen-loc ::= ENUMERATED {
        not-specified (0),
        extracellular (1),
        cytoplasm (2),
        organelle (3),
        nucleus (4),
        membr-cell-cyt (5),
        membr-cell-in (6),
        membr-cell-ext (7),
        membr-outer-peri (8),
        membr-outer-in (9),
        membr-outer-ext (10),
        cellwall-cell (11),
        cellwall-in (12),
        cellwall-ext (13),
```

```
        other(255)
        }

-- ****************************************************************
-- Specific start and end locations for an interaction
--
-- Field description for BIND-spec-loc
-- ***********************************
-- start = specific location where this interaction takes place
-- end = specific location where this interaction ends
-- ****************************************************************

BIND-spec-loc-set ::= SEQUENCE {
        start BIND-spec-loc,
        end BIND-spec-loc OPTIONAL
        }


-- ************************************************************
-- Specific location of interaction with respect to a cell
--
-- Field description for BIND-spec-loc
-- ***********************************
-- location = text specific location
-- descr = text location further description
-- cell-type = text cell type
-- ************************************************************

BIND-spec-loc ::= SEQUENCE {
        location VisibleString,
        descr VisibleString OPTIONAL,
        cell-type VisibleString OPTIONAL
        }


-- ****************************
-- * Interaction conditions *
-- ****************************

-- ***********************************
-- A list of experimental conditions.
-- ***********************************

BIND-condition-set ::= SEQUENCE OF BIND-conditions


-- ************************************************************
-- An experimental condition that has been used to observe
-- this interation.  Interaction must be reproduceable
-- using this information.
--
-- Field description for BIND-conditions
-- *************************************
-- conditions = list of possible experimental conditions
-- system = experimental system used
-- descr = text description
-- source = empirical evidence
```

```
-- ********************************************************

BIND-conditions ::= SEQUENCE {
        conditions ENUMERATED {
                in-vitro(0),
                in-vivo(1),
                other(255)
                },
        system VisibleString OPTIONAL,
        descr VisibleString OPTIONAL,
        source BIND-pub-set OPTIONAL
        }


-- ***********************************
-- * Interaction conserved sequence *
-- ***********************************

-- ********************************************************************
-- Conserved sequence set
--
-- Only relevant for biological sequences
-- Derived from alignment information.
--
-- Field description for BIND-cons-seq-set
-- ****************************************
-- a = conserved sequence of 'a'
-- b = as above, for 'b'
-- ********************************************************************


BIND-cons-seq-set ::= SEQUENCE {
        a BIND-conserved-seq OPTIONAL,
        b BIND-conserved-seq OPTIONAL
        }


-- ********************************************************************
-- Conserved sequence
--
-- Field description for BIND-conserved-seq
-- ****************************************
-- seq-el = these sequence elements have been shown to be conserved
-- descr = further text description
-- source = empirical evidence
-- ********************************************************************

BIND-conserved-seq ::= SEQUENCE {
        seq-el Seq-loc,
        descr VisibleString OPTIONAL,
        source BIND-pub-set OPTIONAL
        }


-- *********************************
-- * Interaction chemical action *
-- *********************************

-- ************************************************
```

```
-- A set of chemical actions.
--
-- Field description for BIND-action
-- **********************************
-- max-iaid = the highest iaid used so far in this set
-- actions = set of BIND-action objects
-- ***************************************************

BIND-action-set ::= SEQUENCE {
        max-iaid Internal-Action-id,
        actions SEQUENCE OF BIND-action
        }

-- ***************************************************************
-- A chemical action.
--
-- Field description for BIND-action
-- **********************************
-- iaid = iid-internal action id
--        (unique for each action in a BIND-action sequence in
--        an interaction)
-- descr = further text description
-- a-on-b = TRUE if 'a' is acting on 'b', FALSE for opposite
-- action = choice of none, add, remove, change, cut, other
--          (contains an object and a location for each action)
-- kinetics = chemical action kinetics
-- source = empirical evidence
-- ***************************************************************

BIND-action ::= SEQUENCE {
        iaid Internal-Action-id,
        descr VisibleString OPTIONAL,
        a-on-b BOOLEAN DEFAULT TRUE,
        action CHOICE {
                none NULL,
                add BIND-action-object,
                remove BIND-action-object,
                change BIND-action-object,
                cut BIND-action-object,
                other BIND-action-object
                },
        kinetics BIND-kinetics OPTIONAL,
        source BIND-pub-set OPTIONAL
        }

Internal-Action-id ::= INTEGER

-- ***************************************************
-- Object used by BIND-action.
--
-- Field description for BIND-action-object
-- ****************************************
-- what = object that is being acted upon (for none, add,
--        remove, cut, other)
-- to = object that this was changed to (for change only)
-- where = location of action
-- descr = optional text description of this object
```

```
-- ********************************************************

BIND-action-object ::= SEQUENCE {
        what BIND-object,
        to BIND-object OPTIONAL,
        where Chem-graph-pntrs,
        descr VisibleString OPTIONAL
        }

-- ********************************************************
-- Chemical kinetics
--
-- Field description for BIND-kinetics
-- ***********************************
-- descr = optional text description of this object
-- kd = dissociation constant of interaction
-- km = Michaelis-Menten constant
-- vmax = max. velocity of reaction
-- conc-a = concentration of 'a' object
-- conc-b = as above, for 'b'
-- temp = temperature of the interaction system (observed)
-- pH = pH of the interaction system
-- half-life-a = 1/2 life for 'a'
-- half-life-b = 1/2 life for 'b'
-- buffer = buffer text description
-- other = any other kinetic related values (e.g. k1, k2...)
-- source = empirical evidence
-- ********************************************************

BIND-kinetics ::= SEQUENCE {
        descr VisibleString OPTIONAL,
        kd RealVal-Units OPTIONAL,
        km RealVal-Units OPTIONAL,
        vmax RealVal-Units OPTIONAL,
        conc-a RealVal-Units OPTIONAL,
        conc-b RealVal-Units OPTIONAL,
        temp RealVal-Units OPTIONAL,
        ph RealVal-Units OPTIONAL,
        half-life-a RealVal-Units OPTIONAL,
        half-life-b RealVal-Units OPTIONAL,
        buffer VisibleString OPTIONAL,
        other SEQUENCE OF BIND-kinetics-other OPTIONAL,
        source BIND-pub-set OPTIONAL
        }

BIND-kinetics-other ::= SEQUENCE {
        descr VisibleString,
        value RealVal-Units
        }

-- ****************************************************************
-- A Real Number
--
-- Field description for RealVal
-- *****************************
-- Basic scientific notation
-- scaled-real-value * 10^(scale-factor)
```

```
-- units = string value of the units involved, e.g. ml, M, etc.
-- ************************************************************


RealVal-Units ::= SEQUENCE {
        scale-factor                INTEGER,
        scaled-real-value  REAL,
        units VisibleStringOPTIONAL
        }


-- ************************************
-- * Interaction object chemical state *
-- ************************************

-- ********************************************************************************************
-- Chemical state and required chemical state for objects 'a' and 'b'
--
-- Field description for BIND-state-descr
-- ************************************
-- a = list of possible activity states for 'a'
-- a-required-state = the state that 'a' is required to assume before interaction
--                    takes place.
-- b = list of possible activity states for 'b'
-- b-required-state = the state that 'b' is required to assume before interaction
--                    takes place.
-- ********************************************************************************************


BIND-state-descr ::= SEQUENCE {
        a BIND-state-set OPTIONAL,
        a-required-state BIND-required-state OPTIONAL,
        b BIND-state-set OPTIONAL,
        b-required-state BIND-required-state OPTIONAL
        }


-- ******************************************************
-- A set of chemical states
--
-- Field description for BIND-state-set
-- ************************************
-- max-isid = highest Internal-State-id used in this set
-- states = list of possible chemical states
-- ******************************************************


BIND-state-set ::= SEQUENCE {
        max-isid Internal-State-id,
        states SEQUENCE OF BIND-state
        }

Internal-State-id ::= INTEGER
```

```
-- ************************************************************************
-- A chemical state
--
-- Field description for BIND-state
-- ********************************
-- isid = Internal-State-id (unique for each state in a BIND-state-set)
-- activity = activity of molecule
-- activity-level = integer level of activity, baseline activity=0
-- cause = sequence of actions that caused current level of activity
-- descr = further text description
-- source = empirical evidence for this state
-- ************************************************************************


BIND-state ::= SEQUENCE {
        isid Internal-State-id,
        activity ENUMERATED {
                none (0),
                active (1),
                inactive (2)
                },
        activity-level INTEGER OPTIONAL,
        cause SEQUENCE OF BIND-state-cause OPTIONAL,
        descr VisibleString OPTIONAL,
        source BIND-pub-set OPTIONAL
        }


-- ****************************************************************
-- Cause of a chemical state
--
-- Uniquely locates a chemical action by iid then by int-ac-id.
--
-- Field description for BIND-state-cause
-- **************************************
-- from-iid = chemical action is from this iid
-- cause = chemical action id number that caused this activity
-- ****************************************************************


BIND-state-cause ::= SEQUENCE {
        from-iid Interaction-id,
        cause Internal-Action-id
        }


-- ************************************************
-- A required chemical state
--
-- Field description for BIND-required-state
-- *****************************************
-- isid = Internal-State-id of the required state
-- source = empirical evidence
-- ************************************************


BIND-required-state ::= SEQUENCE {
        isid Internal-State-id,
        source BIND-pub-set OPTIONAL
        }
```

```
-- **********************
-- * Molecular-Complex *
-- **********************

-- ****************************
-- A set of Molecular Complexes
-- ****************************

BIND-complex-set ::= SEQUENCE OF BIND-Molecular-Complex


-- ************************************************************************
-- A molecular complex record
--
-- Field description for BIND-molecular-complex
-- ******************************************
-- date = date of record entry
-- updates = a list of updates for the record
-- mcid = molecular complex accession number.
-- descr = text description of interaction
-- sub-num = total number of sub-units in this complex
-- sub-units = list of pointers to the actual sub-units
-- interaction-order = the order of interactions that take place to form
--                     this complex.
-- complex-assembly = a chemical graph of the interaction complex
--                    (with molecules as nodes)
-- source = empirical evidence references
-- sub = contact information of the submitter and general information about
--       this submission
-- priv = TRUE if this complex is private
-- ************************************************************************


BIND-Molecular-Complex ::= SEQUENCE {
        date Date,
        updates BIND-update-set OPTIONAL,
        mcid Molecular-Complex-id,
        descr VisibleString OPTIONAL,
        sub-num INTEGER,
        sub-units SEQUENCE OF BIND-object,
        interaction-order SEQUENCE OF Interaction-id,
        complex-assembly Biostruc-graph OPTIONAL,
        source BIND-pub-set,
        sub Submit-block,
        priv BOOLEAN DEFAULT FALSE
        }


-- **********************************
-- * Biomolecular chemical pathway *
-- **********************************

-- ****************************
-- A set of Molecular Complexes
-- ****************************

BIND-pathway-set ::= SEQUENCE OF BIND-Pathway
```

```
-- ********************************************************************
-- A pathway record.
--
-- Field description for BIND-pathway
-- ***********************************
-- date = date of record entry
-- updates = a list of updates for the record
-- pid = pathway accession number
-- pathway = a collection of interactions and signal modification objects
-- descr = descriptors for a pathway
-- source = empirical evidence references
-- sub = contact information of the submitter and general information about
--          this submission
-- priv = TRUE if this pathway is private
-- ********************************************************************


BIND-Pathway ::= SEQUENCE {
        date Date,
        updates BIND-update-set OPTIONAL,
        pid Pathway-id,
        pathway SEQUENCE OF BIND-pathway-object,
        descr BIND-path-descr,
        source BIND-pub-set,
        sub Submit-block,
        priv BOOLEAN DEFAULT FALSE
        }


-- ************************************************************
-- One node in a pathway graph
--
-- Field description for BIND-pathway-object
-- *****************************************
-- iid = interaction id reference
-- signal = pathway signal change mediated by this iid.
-- ************************************************************


BIND-pathway-object ::= SEQUENCE {
        iid Interaction-id,
        signal BIND-delta-signal
        }


-- ************************************************************
-- A chemical signal change
--
-- Field description for BIND-delta-signal
-- ***************************************
-- action = signal modification
-- factor = the factor of the amplification or the repression
-- descr = further text description
-- ************************************************************


BIND-delta-signal ::= SEQUENCE {
        action ENUMERATED {
                none (0),
                amplify (1),
```

```
                repress (2),
                auto-amplify-a(3),
                auto-repress-a(4),
                other (255)
                },
        factor RealVal-Units OPTIONAL,
        descr VisibleString OPTIONAL
        }


-- ************************************************************************
-- Pathway description
--
-- Field description for BIND-path-descr
-- *************************************
-- descr = text description of pathway
-- cell-cycle = if applicable, stage of a cell cycle that this pathway
--              is in effect
-- developmental-stage = developmental stage of an organism,
--              if applicable, that this pathway is in effect
-- ************************************************************************


BIND-path-descr ::= SEQUENCE {
        descr VisibleString OPTIONAL,
        cell-cycle BIND-cellstage OPTIONAL,
        developmental-stage BIND-devstage OPTIONAL
        }


-- ************************************************************************
-- Cell cycle stage
--
-- Field description for BIND-cellstage
-- ************************************
-- phase = phase of cell cycle
-- descr = text description of cell stage
-- ************************************************************************

BIND-cellstage ::= SEQUENCE {
        phase ENUMERATED {
                none (0),
                constitutive (1),
                interphase (2),
                division (3),
                g1 (4),
                s (5),
                g2 (6),
                mitosis (7),
                prophase (8),
                prometaphase (9),
                metaphase (10),
                anaphase (11),
                telophase (12),
                cytokinesis (13),
                meiosis (14),
                prophase1 (15),
                leptotene (16),
                zygotene (17),
                pachytene (18),
```

```
                diplotene (19),
                diakinesis (20),
                metaphase1(21),
                anaphase1 (22),
                telophase1 (23),
                meiotic-cytokinesis (24),
                prophase2 (25),
                metaphase2 (26),
                anaphase2 (27),
                telophase2 (28),
                meiotic-cytokinesis2 (29),
                other (255)
                },
        descr VisibleString OPTIONAL
        }


-- ************************************************************************
-- Organism developmental stage
--
-- Field description for BIND-devstage
-- ************************************
-- stage = text description of developmental stage
-- ************************************************************************


BIND-devstage ::= SEQUENCE {
        stage VisibleString
        }


-- *********************************
-- * GI/DI cross reference record *
-- *********************************

-- ************************************************************************
-- Cross reference for gi/di searching
--
-- Field description for BIND-Cross-Ref
-- ************************************
-- gi = gi number
-- di = di number
-- iids = list of interactions that this gi is involved in
-- pids = list of pathways that this gi is involved in
-- mcids = list of molecular complexes that this gi is involved in
-- ************************************************************************


BIND-Cross-Ref ::= SEQUENCE {
        gi Geninfo-id DEFAULT 0,
        di Domain-id DEFAULT 0,
        iids SEQUENCE OF Interaction-id,
        pids SEQUENCE OF Pathway-id OPTIONAL,
        mcids SEQUENCE OF Molecular-Complex-id OPTIONAL
        }


-- ************************************
-- * PMID/MUID cross reference record *
-- ************************************
```

```
-- ***************************************************************************
-- Cross reference for pmid/muid searching
--
-- Field description for BIND-Pub-Cross-Ref
-- *************************************
-- uid = muid or pmid
-- iids = list of interactions with this publication as a reference
-- pids = list of pathways with this publication as a reference
-- mcids = list of molecular complexes with this publication as a reference
-- ***************************************************************************


BIND-Pub-Cross-Ref ::= SEQUENCE {
        uid INTEGER,
        iids SEQUENCE OF Interaction-id,
        pids SEQUENCE OF Pathway-id OPTIONAL,
        mcids SEQUENCE OF Molecular-Complex-id OPTIONAL
        }

END
```

APPENDIX B

```
--$Revision: 1.1 $
-- ********************************************************************************
--
-- Biomolecular Interaction Network Database (BIND)
-- Data Specification
--
-- Interaction, Molecular Complex, Biological Pathway Data Structures
--
--
-- Authors: Gary D. Bader,      Christopher W.V. Hogue
--        bader@mshri.on.ca    hogue@mshri.on.ca
--
-- --
-- Hogue Lab - University of Toronto Biochemistry Department and the
-- Samuel Lunenfeld Research Institute, Mount Sinai Hospital
-- http://bioinfo.mshri.on.ca  hogue@mshri.on.ca
--
-- REVISIONS
-- Revision 0.1 - Oct. 21, 1998
-- Revision 0.5 - Feb. 2,  1999 (BIND web based data entry prototype)
-- Revision 0.6 - Feb. 26, 1999 (Feedback from Biophysical Soc. Conf.)
-- Revision 0.8 - May  3,  1999
-- Revision 0.9 - May  31, 1999
-- Revision 1.0 - June 7,  1999 (comments only added to 0.9)
-- Revision 1.1 - Some changes and additions
--         Removed iid from BIND-object
--         Changed Pub to pub-set in database-site
--         Moved BIND-cellstage object definition higher in specification (aesthetic change)
--         Removed OPTIONAL from activity-level in BIND-state
--         Added 'not-specified' to BIND-Submitter/subtype and removed the OPTIONAL,
--         Added an OPTIONAL BIND-mol-sub-num object to BIND-mol-object
--          to account for more complicated molecular complexes
--         Added OPTIONAL BIND-condition-dependency to BIND-loc-site
--         Added boolean intramolecular field to BIND-descr
--         Added 'none' to BIND-action-type to allow e.g. kinetics data to be stored with no action
--         Added enzyme activity amplification factor to listed kinetics values
--         Added structured address fields to BIND-Contact-info object
--         Added pathological-state to BIND-path-descr
--         Added author list to Interaction, Complex and Pathway record
--         Added more options to BIND-gen-place enumerated type
--         Added 'break' as choice in BIND-action-type object
--         Added list of synonyms for short-label in object
--         Added list of complex exclusive interactions in molecular complex record
--         Added in-situ to BIND-condition general field as a choice
--         Added optional BioSource to BIND-chemsource for natural products
--         Added simple 'active' choice to BIND-state activity-level to allow description of active/inactive state
--         Removed sub fields from IID,MCID,PID cross reference records.  This info clutters the cross-reference
record
--          and will make a cross-ref system less efficient and reduces privacy of submitter information.  Rather
save
--          submitter information in another database that the specification does not impose structure on.
```

```
--          Added generic BIND-direction object to describe four cases a-a,a-b,b-a,b-b
--          Changed a-on-b field in BIND-action and a-to-b field in BIND-signal to BIND-direction.  In BIND-
action
--           this will account for autophosphorylation and similar self changing events
--          Removed auto-amplify and auto-repress as choices in BIND-signal
--          Added optional list of accession numbers to BIND-submit
--          Added better description of submission tool to BIND-Submitter
--          Added sub-unit fields to BIND-loc-site, BIND-action and BIND-state so a sub-unit in a complex
--           can be specified.
--          Added optional text description to BIND-required-state
--          Added optional text description to BIND-loc-site
--          Added Bioseq in experimental sequence field in BIND-condition
--
-- ftp://bioinfo.mshri.on.ca/pub/BIND/Spec/bind.asn for the latest revision.
--
--
-- NOTE:  This specification is in a variant of ASN.1 1990 that may not
--       be compatible with newer ASN.1 tools.  This specification also
--       depends on v6.1 public domain specifications available from the
--       U.S. National Center for Biotechnology Information (NCBI)
--       ftp://ncbi.nlm.nih.gov/toolbox/ncbi_tools/
--       http://www.ncbi.nlm.nih.gov/Toolbox/
--
--
--
-- ********************************************************************************

BIND DEFINITIONS ::=
BEGIN

EXPORTS BIND-Submit,
          BIND-Interaction, BIND-Interaction-set,
          BIND-Pathway, BIND-Pathway-set,
          BIND-Molecular-Complex, BIND-Complex-set;

IMPORTS Date, Int-fuzz FROM NCBI-General
          Author FROM NCBI-Biblio
          Bioseq FROM NCBI-Sequence
          Pub FROM NCBI-Pub
          BioSource FROM NCBI-BioSource
          Seq-loc, Seq-id FROM NCBI-Seqloc
          Biostruc FROM MMDB
          Biostruc-feature-set FROM MMDB-Features;


-- <!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!>
-- ****************************************
-- * Data Submission and Database exchange *
-- ****************************************

-- ********************************************************************************
-- General data exchange
--
-- This object is used to submit all information to BIND.
--
```

```
-- Field description for BIND-Submit
-- **********************************
-- date = date of creation of this set
-- database = description of database where this data originated
-- sub = person who is responsible for this data submission
-- sub-id = BIND submit ID for this submission
-- acc-nums = list of BIND accession numbers that this submission contains
-- interactions = a collection of interaction records
-- complexes = a collection of molecular complex records
-- pathways = a collection of pathway records
-- ********************************************************************************


BIND-Submit ::= SEQUENCE {
        date Date,
        database BIND-Database-site OPTIONAL,
        sub BIND-Submitter,
        sub-id BIND-Submit-id OPTIONAL,
        acc-nums SEQUENCE OF BIND-accession-number OPTIONAL,
        interactions BIND-Interaction-set OPTIONAL,
        complexes BIND-Complex-set OPTIONAL,
        pathways BIND-Pathway-set OPTIONAL
        }

BIND-Submit-id ::= INTEGER

BIND-accession-number ::= CHOICE {
        interaction Interaction-id,
        complex Molecular-Complex-id,
        pathway Pathway-id
        }


-- *************************
-- * Database description *
-- *************************

-- ********************************************************************************
-- Description of a database site
--
-- Field description for BIND-Database-Site
-- ****************************************
-- descr = text description of this database
--         (e.g. C. elegans interaction database)
-- country = country where this database is based
-- homepage-url = Internet Universal Resource Locator for the database web site
--                (e.g. http://bioinfo.mshri.on.ca)
-- reference = a Medline reference for this database
-- ********************************************************************************


BIND-Database-site ::= SEQUENCE {
        descr VisibleString,
        country VisibleString,
        homepage-url VisibleString OPTIONAL,
        reference BIND-pub-set OPTIONAL
        }
```

```
-- **************
-- * Submitter *
-- **************

-- ********************************************************************************
-- Description of a submitter (Adaptation of NCBI Submit-Block)
--
-- Field description for BIND-Submitter
-- ************************************
-- contact = submitter contact information
-- hup = hold this submission until published
-- subtype = submission type
-- tool = tool used to submit record (e.g. BIND Web Data Entry version 1.0)
-- ********************************************************************************


BIND-Submitter ::= SEQUENCE {
        contact BIND-Contact-info,
        hup BOOLEAN DEFAULT FALSE,
        subtype ENUMERATED {
                not-specified (0),
                new (1),            -- new data
                update (2),         -- update by author
                revision (3),       -- 3rd party (non-author) update
                other (255) },
        tool BIND-Submission-tool OPTIONAL
        }

BIND-Submission-tool SEQUENCE {
        name VisibleString,
        version VisibleString,
        descr VisibleString OPTIONAL
        }

-- ************************
-- * Contact Information *
-- ************************

-- ********************************************************************************
-- Contact information (Adaptation of NCBI Contact-info)
--
-- Field description for BIND-Contact-info
-- ***************************************
-- first-name = First name of submitter
-- middle-initial = Middle initial of submitter
-- last-name = Last name of submitter
-- address = Street address of submitter
-- room = Room number
-- dept = Department
-- institute = Institute if this is different than organization
--          (e.g. research institute)
-- organization = Organization (e.g. University of Toronto)
-- city = City
-- pcode = Zip or postal code
-- country = Country
-- phone = Phone number (with area code)
```

```
-- fax = Fax number (with area code)
-- email = E-mail address
-- userid = User ID number
-- other = any other contact information
-- ****************************************************************************

BIND-Contact-info ::= SEQUENCE {
        first-name VisibleString OPTIONAL,
        middle-initial VisibleString OPTIONAL,
        last-name VisibleString OPTIONAL,
        address SEQUENCE OF VisibleString OPTIONAL,
        room VisibleString OPTIONAL,
        dept VisibleString OPTIONAL,
        institute VisibleString OPTIONAL,
        organization OPTIONAL,
        city VisibleString OPTIONAL,
        pcode VisibleString OPTIONAL,
        country VisibleString OPTIONAL,
        phone VisibleString OPTIONAL,
        fax VisibleString OPTIONAL,
        email VisibleString OPTIONAL,
        userid INTEGER OPTIONAL,
        other SEQUENCE OF VisibleString OPTIONAL
        }


-- ****************
-- * Publications *
-- ****************

-- ****************************************************************************
-- A set of publications
--
-- Field description for BIND-pub-set
-- ***********************************
-- disputed = TRUE if a BIND-pub-object in this set contains a dispute flag
-- pubs = a sequence of BIND-pub-objects
-- ****************************************************************************

BIND-pub-set ::= SEQUENCE {
        disputed BOOLEAN DEFAULT FALSE,
        pubs SEQUENCE OF BIND-pub-object
        }


-- ****************************************************************************
-- A publication
--
-- Field description for BIND-pub-object
-- *************************************
-- descr = text description of this object
-- opinion = does this publication support or dispute the data?
-- pub = full NCBI publication reference

-- ****************************************************************************
```

```
BIND-pub-object ::= SEQUENCE {
        descr VisibleString OPTIONAL,
        opinion ENUMERATED {
                none (0),
                support (1),
                dispute (2)
                },
        pub Pub
        }


-- ******************
-- * Record Update *
-- ******************

-- ********************************************************************************
-- An update for a record
--
-- Field description for BIND-update
-- *********************************
-- date = date of this update
-- descr = text description of update (this can store any update information
--       up to the entire previous version of the record in ASN.1)
-- ********************************************************************************


BIND-update-object ::= SEQUENCE {
        date Date,
        descr VisibleString
        }


-- ********************************************************************************
-- Cell cycle stage
--
-- Field description for BIND-cellstage
-- ************************************
-- phase = phase of cell cycle
-- descr = text description of cell stage (e.g. if 'other' is specified)
-- ********************************************************************************


BIND-cellstage ::= SEQUENCE {
        phase INTEGER {
                not-specified (0),
                constitutive (1),
                interphase (2),
                division (3),
                g1 (4),
                s (5),
                g2 (6),
                mitosis (7),
                prophase (8),
                prometaphase (9),
                metaphase (10),
```

```
                anaphase (11),
                telophase (12),
                cytokinesis (13),
                meiosis (14),
                prophase1 (15),
                leptotene (16),
                zygotene (17),
                pachytene (18),
                diplotene (19),
                diakinesis (20),
                metaphase1(21),
                anaphase1 (22),
                telophase1 (23),
                meiotic-cytokinesis (24),
                prophase2 (25),
                metaphase2 (26),
                anaphase2 (27),
                telophase2 (28),
                meiotic-cytokinesis2 (29),
                other (255)
                },
        descr VisibleString OPTIONAL
        }


-- ********************************************************************************
-- A Real Number
--
-- Field description for RealVal-Units
-- ***********************************
-- scaled-integer-value * 10^(scale-factor)
-- units = string value of the units involved (e.g. ml, M, etc.)
-- ********************************************************************************


RealVal-Units ::= SEQUENCE {
        scale-factor            INTEGER,
        scaled-integer-value    INTEGER,
        units VisibleStringOPTIONAL
        }


-- <!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!>
-- ***************
-- * Interaction *
-- ***************

-- ********************************************************************************
-- A set of interactions
--
-- Field description for BIND-Interaction-set
-- ******************************************
-- date = date this set of records was collected
-- database = name and description of database that this set originates
-- interactions = set of interaction records
-- ********************************************************************************
```

```
BIND-Interaction-set ::= SEQUENCE {
        date Date OPTIONAL,
        database BIND-Database-site OPTIONAL,
        interactions SEQUENCE OF BIND-Interaction
        }

-- ********************************************************************************
-- A BIND-Interaction record can store all of the details of an interaction
-- between any two molecules (or atoms).
--
-- Field description for BIND-Interaction
-- ************************************
-- date = date of record entry
-- updates = a list of updates for this record
-- iid = interaction accession number
-- a = molecule 'a' interacts with...
-- b = molecule 'b'
-- descr = description of interaction
-- source = empirical evidence references (publications)
-- authors = person(s) who authored this record.
-- priv = TRUE if this interaction is private
-- ********************************************************************************
-- NOTE: In the context of this data specification, the 'priv' flag means:
--         -Do not export this record.
--         -In a public database, this record is not available to be publicly
--          retrieved.
--         -In a private database, this record can be retrieved, but it will
--          not be exported.
-- ********************************************************************************

BIND-Interaction ::= SEQUENCE {
        date Date,
        updates SEQUENCE OF BIND-update-object OPTIONAL,
        iid Interaction-id,
        a BIND-object,
        b BIND-object,
        descr BIND-descr,
        source BIND-pub-set,
        authors SEQUENCE OF Author OPTIONAL,
        priv BOOLEAN DEFAULT FALSE
        }

Interaction-id ::= INTEGER

-- ************************
-- * Biomolecular Object *
-- ************************

-- ********************************************************************************
-- Any chemical object
--
-- Field description for BIND-object
```

```
-- ********************************
-- short-label = short label of this object (e.g. ATP, S4, HSP70)
-- short-label-syn = list of short-label synonyms for this object
-- id = the type of chemical object and a pointer to a record in a database
--          of the object type (e.g. protein database)
--          Choice of: not-specified, protein, DNA, RNA, small-molecule (any other type of
--                    chemical compound), interaction, molecular complex
-- origin = material source (biological or chemical origin)
-- cell-stage = description of cell cycle stages this object is specific to
-- seq = space for sequence, if it is not in a public database
--          ALSO, this can be a consensus sequence for binding of this object
--          (e.g. transcription factor binding to DNA)
-- struc = space for complete structure, if not in public database
--          (This should not be used to store a structure that is already in
--          the MMDB)
-- descr = text description of this object
-- ************************************************************************

BIND-object ::= SEQUENCE {
          short-label VisibleString,
          short-label-syn SEQUENCE OF VisibleString OPTIONAL,
          id BIND-object-type-id,
          origin BIND-object-origin,
          cell-stage SEQUENCE OF BIND-cellstage OPTIONAL,
          seq Bioseq OPTIONAL,
          struc Biostruc OPTIONAL,
          descr VisibleString OPTIONAL
          }

BIND-object-type-id ::= CHOICE {
          not-specified NULL,
          protein BIND-id,
          dna BIND-id,
          rna BIND-id,
          small-molecule BIND-small-molecule-id,
          complex Molecular-Complex-id
          }

BIND-object-origin ::= CHOICE {
          not-specified NULL,
          org BioSource,
          chem BIND-chemsource
          }

-- ************************************************************************
-- Summary description of a chemical compound
--
-- Field description for BIND-chemsource
-- ************************************
-- names = chemical compound name and any synonyms
-- smiles-string = standard smiles-string for this compound
--  References for SMILES language:
--   D. Weininger, SMILES, a Chemical Language and Information System.
--   1. Introduction to Methodology and Encoding Rules,
--   J. Chem. Inf. Comput. Sci. 1988, 28, 31-36.
--  Web sites:
```

```
--    http://www.daylight.com/dayhtml/smiles/smiles-intro.html
--    http://www2.ccc.uni-erlangen.de/services/smiles.html
-- molecular-weight = molecular weight of this compound in g/mol
-- chemical-formula = chemical formula of the compound (e.g.C3H7NO2)
-- cas-number = Chemical Abstracts Service (http://www.cas.org/)
--                  database number for this compound (e.g. 56-41-7)
-- nat-prod = biological source information if this is a natural product
-- ****************************************************************************


BIND-chemsource ::= SEQUENCE {
        names SET OF VisibleString,
        smiles-string VisibleString OPTIONAL,
        chemical-formula VisibleString OPTIONAL,
        molecular-weight RealVal-Units OPTIONAL,
        cas-number VisibleString OPTIONAL,
        nat-prod BioSource OPTIONAL
        }


-- ***************
-- * Identifiers *
-- ***************

-- ****************************************************************************
-- General sequence or domain identifier
--
-- Field description for BIND-id
-- ****************************
-- gi = NCBI integer accession number (optional only for sequence data with
--        no NCBI database identifier).
--        NOTE: gi is stored so that a BIND-object refers to a constant sequence
--        molecule.  This is necessary to maintain data integrity of Seq-loc's
--        also stored in the BIND database.
-- di = domain accession number (from the domain split database)
-- other = open field for other possible NCBI defined pointers
--          (if possible, equivalent GenBank accession number to this
--      gi should be stored here as well)
--
-- NOTE: there is a field for gi in a Seq-id, but it should not be used
-- in this object
-- ****************************************************************************


BIND-id ::= SEQUENCE {
        gi Geninfo-id OPTIONAL,
        di Domain-id OPTIONAL,
        other SET OF Seq-id OPTIONAL
        }

Geninfo-id ::= INTEGER

Domain-id ::= INTEGER


-- ****************************************************************************
-- Pointer to a small molecule database
```

```
--
-- Field description for BIND-small-molecule-id
-- ************************************
-- internal = id number of an internally kept record of a chemical compound
-- other-db = generic pointer to any other database (e.g. Japanese LIGAND db)
--          Contains the name of the database, an integer pointer and/or a string
--          pointer.
-- ************************************************************************

BIND-small-molecule-id ::= CHOICE {
          internal Internal-small-molecule-id,
          other-db BIND-other-db
          }

Internal-small-molecule-id ::= INTEGER

BIND-other-db ::= SEQUENCE {
          dbname VisibleString,
          intp INTEGER OPTIONAL,
          strp VisibleString OPTIONAL
          }

-- ************************************************
-- * Interaction Description (in BIND-Interaction) *
-- ************************************************

-- ************************************************************************
-- Full description of an interaction
--
-- Field description for BIND-descr
-- ********************************
-- simple-descr = text description of this interaction
-- place = description of cellular place of interaction
-- cond = binding conditions/experimental conditions
-- cons = conserved sequence comment
-- binding-sites = location of binding sites on object A and B
-- action = list of chemical actions that can occur in this interaction
-- state = list of chemical states of 'a' and 'b' as well as required
--          state for interaction to occur
-- intramolecular = only relevant if 'a' and 'b' refer to the same molecule.
--          TRUE if the interaction is intramolecular
-- ************************************************************************

BIND-descr ::= SEQUENCE {
          simple-descr VisibleString OPTIONAL,
          place SEQUENCE OF BIND-place OPTIONAL,
          cond BIND-condition-set OPTIONAL,
          cons BIND-cons-seq-set OPTIONAL,
          binding-sites BIND-loc OPTIONAL,
          action BIND-action-set OPTIONAL,
          state BIND-state-descr OPTIONAL,
          intramolecular BOOLEAN DEFAULT FALSE
          }
```

```
-- ********************************************
-- * Cellular Interaction Place (in BIND-descr) *
-- ********************************************

-- ************************************************************************************
-- Location of interaction with respect to the cell
--
-- Field description for BIND-place
-- ********************************
-- bpid = internal BIND place ID number
-- gen-place = general cellular locations where this interaction takes place
--         (computer readable)
-- spec-place = specific text locations of the interaction
--         (human readable)
-- source = empirical evidence references (publications)
-- descr = text description (e.g. method of finding interaction place)
-- ************************************************************************************


BIND-place ::= SEQUENCE {
        bpid BIND-place-id OPTIONAL,
        gen-place BIND-gen-place-set,
        spec-place BIND-spec-place-set OPTIONAL,
        source BIND-pub-set OPTIONAL,
        descr VisibleString OPTIONAL
        }

BIND-place-id ::= INTEGER

-- ************************************************************************************
-- General start and end places for an interaction
--
-- Field description for BIND-gen-place-set
-- ****************************************
-- start = general place in the cell where this interaction takes place
-- end = general place in the cell where this interaction ends
--         (e.g. for translocation)
-- descr = text description (e.g. mechanism of translocation)
-- ************************************************************************************


BIND-gen-place-set ::= SEQUENCE {
        start BIND-gen-place,
        end BIND-gen-place OPTIONAL,
        descr VisibleString OPTIONAL
        }


-- ************************************************************************************
-- General cellular place where this interaction takes place
--
-- This object is meant to be computer readable for e.g. a pathway
-- drawing program. Further cell locations are not enumerated because
-- there are too many in biology.
--
-- Field description for BIND-gen-place
```

```
-- ************************************
-- An enumeration of general cell places
--         not-specified = not-specified
--         extracellular = extracellular
--         cytoplasm = in cytoplasm
--         organelle = in an organelle
--         nucleus = in nucleus
--         membr-cell-cyt = on the cytoplasmic side of the cell membrane
--         membr-cell-in = in the cell membrane
--         membr-cell-ext = on the surface of the cell membrane
--         membr-outer-peri = on the periplasmic side of the outer membrane
--         membr-outer-in = in the outer membrane
--         membr-outer-ext = on the surface of the outer membrane
--         cellwall-cell = on the inside surface of cell wall
--         cellwall-in = in the cell wall
--         cellwall-ext = on the outside surface of the cell wall
--         other = other location - see text description in BIND-gen-loc-set
-- ************************************************************************
BIND-gen-place ::= ENUMERATED {
        not-specified (0),
        extracellular (1),
        cytoplasm (2),
        organelle (3),
        nucleus (4),
        membr-cell-cyt (5),
        membr-cell-in (6),
        membr-cell-ext (7),
        membr-outer-peri (8),
        membr-outer-in (9),
        membr-outer-ext (10),
        cellwall-cell (11),
        cellwall-in (12),
        cellwall-ext (13),
        nuclear-envelope (14),
        perinuclear-space (15),
        nuc-inner-membr (16),
        nuc-outer-membr (17),
        nucleolus (18),
        chromatin (19),
        er (20),
        smooth-er (21),
        rough-er (22),
        golgi (23),
        cis-golgi (24),
        trans-golgi (25),
        vacuole (26),
        lysosome (27),
        peroxisome (28),
        mitochondrion (29),
        mito-outer-membr (30),
        mito-inner-membr (31),
        mito-christae (32),
        mito-matrix (33),
        chloroplast (34),
        chlor-inner-membr (35),
        chlor-outer-membr (36),
        thylakoid (37),
        grana (38),
```

```
            stroma (39),
            centrosome (40),
            centriole (41),
            other(255)
            }

BIND-gen-place-expanded ::= CHOICE {
            not-specified NULL,
            extracellular NULL,
            cytoplasm NULL,
            cell-wall BIND-gen-place-membr-descr,
            outer-membrane BIND-gen-place-membr-descr,
            cytoplasmic-membrane BIND-gen-place-membr-descr,
            organelle-unknown BIND-gen-place-membr-descr,
            organelle-other BIND-gen-place-membr-descr,
            nucleus BIND-gen-place-membr-descr,
            er-general BIND-gen-place-membr-descr,
            er-smooth BIND-gen-place-membr-descr,
            er-rough BIND-gen-place-membr-descr,
            golgi BIND-gen-place-membr-descr,
            cis-golgi BIND-gen-place-membr-descr,
            trans-golgi BIND-gen-place-membr-descr,
            vacuole BIND-gen-place-membr-descr,
            lysosome BIND-gen-place-membr-descr,
            peroxisome BIND-gen-place-membr-descr,
            endosome BIND-gen-place-membr-descr,
            mito-general NULL,
            mito-outer-membrane BIND-gen-place-membr-descr,
            mito-inner-membrane BIND-gen-place-membr-descr,
            }

BIND-gen-place-membr-descr ::= ENUMERATED {
            not-specified (0),
            outer-surface (1),
            within (2),
            inner-surface (3),
            lumen (4)
            }


-- ************************************************************************
-- Specific start and end places for an interaction
-- (Human readable)
--
-- Field description for BIND-spec-place
-- ************************************
-- start = specific location where this interaction takes place
--             (e.g. trans golgi, basal membrane, inner mitochondrial space, etc.)
-- end = specific location where this interaction ends
-- ************************************************************************


BIND-spec-place-set ::= SEQUENCE {
            start VisibleString,
            end VisibleString OPTIONAL
            }

-- ****************************************
```

```
-- * Interaction conditions (in BIND-descr) *
-- ******************************************

-- ********************************************************************************
-- A set of experimental conditions.
--
-- Field description for BIND-conditions-set
-- ****************************************
-- max-icid = the highest icid used in this set
-- conditions = set of BIND-condition objects
-- ********************************************************************************

BIND-condition-set ::= SEQUENCE {
        max-icid Internal-conditions-id,
        conditions SEQUENCE OF BIND-condition
        }

-- ********************************************************************************
-- An experimental condition that has been used to observe
-- this interaction.  Interaction conclusion must be reproducible
-- using this information.
--
-- Field description for BIND-conditions
-- *************************************
-- icid = internal condition id
-- general = list of possible general experimental conditions
-- system = experimental system used
-- exp-seq = experimental sequence used if different from actual sequence
--       (e.g. HIS tagged sequence)
-- descr = text description (e.g. if 'other' is specified
--       in conditions or system)
-- source = empirical evidence
-- ********************************************************************************

BIND-condition ::= SEQUENCE {
        icid Internal-conditions-id,
        general ENUMERATED {
                in-vitro(0),
                in-vivo(1),
                in-situ(2),
                in-silico(3),
                other(255)
                },
        system BIND-experimental-system,
        exp-seq Bioseq OPTIONAL,
        descr VisibleString OPTIONAL,
        source BIND-pub-set OPTIONAL
        }

Internal-conditions-id ::= INTEGER

BIND-experimental-system ::= INTEGER {
        not-specified (0),
        alanine-scanning (1),
```

```
        affinity-chromatography (2),
        atomic-force-microscopy (3),
        autoradiography (4),
        competition-binding (5),
        cross-linking (6),
        deuterium-hydrogen-exchange (7),
        electron-microscopy (8),
        electron-spin-resonance (9),
        elisa (10),
        equilibrium-dialysis (11),
        fluorescence-anisotropy (12),
        footprinting (13),
        gel-retardation-assays (14),
        gel-filtration-chromatography (15),
        hybridization (16),
        immunoblotting (17),
        immunoprecipitation (18),
        immunostaining (19),
        interaction-adhesion-assay (20),
        light-scattering (21),
        mass-spectrometry (22),
        membrane-filtration (23),
        monoclonal-antibody-blockade (24),
        nuclear-translocation-assay (25),
        phage-display (26),
        reconstitution (27),
        resonance-energy-transfer (28),
        site-directed-mutagenesis (29),
        sucrose-gradient-sedimentation (30),
        surface-plasmon-resonance-chip (31),
        transient-coexpression (32),
        three-dimensional-structure (33),
        two-hybrid-test (34),
        other (255)
        }

-- ************************************************************
-- * Interaction conserved sequence comment (in BIND-descr) *
-- ************************************************************

-- ********************************************************************************
-- Conserved sequence comment set
--
-- Only relevant for biological sequences.
-- (e.g. Derived from multiple alignment information)
--
-- Field description for BIND-cons-seq-set
-- ****************************************
-- a = conserved sequence comment for molecule 'a'
-- b = conserved sequence comment for molecule 'b'
-- ********************************************************************************

BIND-cons-seq-set ::= SEQUENCE {
        a BIND-conserved-seq OPTIONAL,
        b BIND-conserved-seq OPTIONAL
        }
```

```
-- ****************************************************************************
-- Conserved sequence comment
--
-- Alignment data is not stored here, only the conclusion from it.
--
-- Field description for BIND-conserved-seq
-- ****************************************
-- seq-el = sequence elements that have been shown to be conserved
-- descr = text description (e.g. method of determining conserved sequence)
-- source = empirical evidence
-- ****************************************************************************

BIND-conserved-seq ::= SEQUENCE {
        seq-el Seq-loc,
        descr VisibleString OPTIONAL,
        source BIND-pub-set OPTIONAL
        }

-- ************************************************************************
-- * Binding location on molecules in an interaction (in BIND-descr) *
-- ************************************************************************

-- ****************************************************************************
-- Binding location on a BIND-object
--
-- Field description for BIND-loc
-- ******************************
-- detailed = atomic level detail of interaction sites
-- general = sequence element level description of interaction sites
-- source = empirical evidence
-- ****************************************************************************

BIND-loc ::= SEQUENCE {
        detailed Biostruc OPTIONAL,
        general BIND-loc-gen OPTIONAL,
        source BIND-pub-set OPTIONAL
        }

-- ****************************************************************************
-- General binding location on a BIND-object
--
-- Field description for BIND-loc-gen
-- **********************************
-- a-sites = list of binding sites on object A
-- b-sites = list of binding sites on object B
-- bound = list of sequence elements from A and B that are bound together
-- ****************************************************************************

BIND-loc-gen ::= SEQUENCE {
```

```
        a-sites BIND-loc-site-set OPTIONAL,
        b-sites BIND-loc-site-set OPTIONAL,
        bound SEQUENCE OF BIND-loc-pair OPTIONAL
        }

-- ****************************************************************************
-- A graph describing which sites on A bind to which sites on B
--      BIND-loc-site objects are nodes in the graph
--      BIND-loc-pair objects are edges in the graph
--
-- Field description for BIND-loc-site
-- ***********************************
-- slid = internal ID of this sequence element
-- site = a sequence element (point or interval)
-- condition = this binding site seen only under certain experimental conditions
-- sub-unit = if a or b is a molecular complex, specifies which sub-unit the site is on.
-- descr = description of this binding site
--
-- Field description for BIND-loc-pair
-- ***********************************
-- a-slid = the Seq-loc pointed to by this ID is connected to...
-- b-slid = the Seq-loc pointed to by this ID
-- ****************************************************************************


BIND-loc-site-set ::= SEQUENCE {
        max-slid BIND-Seq-loc-id,
        sites SEQUENCE OF BIND-loc-site
        }

BIND-loc-site ::= SEQUENCE {
        slid BIND-Seq-loc-id,
        site Seq-loc,
        condition BIND-condition-dependency OPTIONAL,
        sub-unit BIND-complex-subunit OPTIONAL,
        descr VisibleString OPTIONAL
        }

BIND-loc-pair ::= SEQUENCE {
        a-slid BIND-Seq-loc-id,
        b-slid BIND-Seq-loc-id
        }

BIND-Seq-loc-id ::= INTEGER


-- ************************************************
-- * Interaction chemical action (in BIND-descr) *
-- ************************************************

-- ****************************************************************************
-- A set of chemical actions
--
-- Chemical actions mediated by a molecule (object 'a' or 'b') in the
-- interaction (a set because a kinase may phosphorylate a protein multiple
```

```
-- times)
--
-- Field description for BIND-action-set
-- ************************************
-- max-iaid = the highest iaid used in this set
-- actions = set of BIND-action objects
-- ********************************************************************************


BIND-action-set ::= SEQUENCE {
        max-iaid Internal-action-id,
        actions SEQUENCE OF BIND-action
        }


-- ********************************************************************************
-- A chemical action
--
-- Field description for BIND-action
-- *********************************
-- iaid = internal action id (unique identifier for this action in a set)
-- descr = text description (e.g. if 'other' is specified for type)
-- direction = direction of chemical action
-- type = type of chemical action
-- result = the product(s) of this chemical action
--  NOTE   this field holds the exact chemical form that is produced, and is
--         used by reference by the next interaction acting on the "product".
--         For a biopolymer this holds the atoms&bonds representation of the
--         molecule.
-- diff = the atomic level detail of differences created by this action
-- signal = more general kinetics, signal transduction
-- kinetics = chemical action kinetics
-- conditions = link to experimental conditions used to find this action,
--                    e.g. if there were multiple experimental conditions stored in
--          this interaction record and this action was only seen using
--          some of them.
-- source = empirical evidence
-- sub-unit-a = if a is a molecular complex, specifies the sub-unit to which
--          this chemical action applies
-- sub-unit-b = if b is a molecular complex, specifies the sub-unit to which
--          this chemical action applies
-- active-site = which site on the acting molecule is performing the chemical action
-- ********************************************************************************


BIND-action ::= SEQUENCE {
        iaid Internal-action-id,
        descr VisibleString OPTIONAL,
        direction BIND-direction,
        type BIND-action-type,
        result SEQUENCE OF BIND-object OPTIONAL,
        diff Biostruc-feature-set OPTIONAL,
        signal BIND-signal OPTIONAL,
        kinetics BIND-kinetics OPTIONAL,
        condition SEQUENCE OF BIND-condition-dependency OPTIONAL,
        source BIND-pub-set OPTIONAL,
```

```
        sub-unit-a BIND-complex-subunit OPTIONAL,
        sub-unit-b BIND-complex-subunit OPTIONAL,
        active-site BIND-active-site OPTIONAL
        }

Internal-action-id ::= INTEGER

BIND-direction ::= ENUMERATED {
        none (0),
        a-to-a (1),
        a-to-b (2),
        b-to-b (3),
        b-to-a (4),
        other (255),
        }

BIND-active-site ::= CHOICE {
        slid BIND-Seq-loc-id,
        site BIND-loc-site-set
        }

-- ********************************************************************************
-- The type of action and object of that action
--
-- Action type          object of that action
-- add                  BIND-object or NULL
-- remove               BIND-object or NULL
-- cut-seq              Seq-loc or NULL
--
-- Field description for BIND-action-type
-- ************************************
-- -not-specified = action is not-specified (unknown)
-- -none = no chemical action, but e.g. kinetics information needs to be stored
--          (action is known to be nothing)
-- -add = add an object (e.g. phosphate) to an object
-- -remove = remove an object (e.g. phosphate) from an object
-- -break = non-sequence cut action - e.g. small molecule hydrolysis
-- -cut-seq = cut a sequence, location may be specified
--          (e.g. restriction enzyme)
-- -change-conformation = a change in conformation of a molecule
--          (e.g. hck protein -> phosphorylation causes conformational change)
-- -change-configuration = a change in configuration of a molecule
--          (e.g. by an epimerase or isomerase)
-- -change-other = another type of change (e.g. metal ion exchange)
-- -other = another action
--
-- Field description for BIND-action-object
-- ****************************************
-- none = no action object
-- object = any BIND-object that is added or removed (e.g. phosphate)
-- location = location where a sequence was cut
-- ********************************************************************************


BIND-action-type ::= CHOICE {
        not-specified NULL,
        none NULL,
```

```
        add BIND-action-object,
        remove BIND-action-object,
        break NULL,
        cut-seq BIND-action-object,
        change-conformation NULL,
        change-configuration NULL,
        change-other NULL,
        other NULL
        }

BIND-action-object ::= CHOICE {
        none NULL,
        object BIND-object,
        location Seq-loc
        }


-- ********************************************************************************
-- A chemical signal description
--
-- A more general notion of kinetics describing signal transduction.
--
-- Field description for BIND-signal
-- *********************************
-- action = signal modification
-- direction = direction of signal
-- factor = the factor of the amplification or the repression
-- descr = text description (e.g. if 'other' is specified)
-- ********************************************************************************


BIND-signal ::= SEQUENCE {
        action ENUMERATED {
                none (0),
                amplify (1),
                repress (2),
                other (255)
                },
        direction BIND-direction,
        factor RealVal-Units OPTIONAL,
        descr VisibleString OPTIONAL
        }


-- ********************************************************************************
-- Chemical kinetics and thermodynamics data
--
-- Field description for BIND-kinetics
-- ***********************************
-- descr = optional text description of this object
-- kd = dissociation constant of interaction
-- km = Michaelis-Menten constant
-- vmax = max. velocity of reaction
-- rxn-order = reaction order
-- conc-a = concentration of 'a'
-- conc-b = concentration of 'b'
-- conc-a-bound = concentration of 'a' that is bound
-- conc-b-bound = concentration of 'b' that is bound
```

```
-- conc-a-unbound = concentration of 'a' that is not bound
-- conc-b-unbound = concentration of 'b' that is not bound
-- enz-activity-amp-factor = scalar amplification factor for enzyme kinetic activity
-- temp = temperature of the interaction system (observed)
-- ph = pH of the interaction system
-- half-life-a = 1/2 life for 'a'
-- half-life-b = 1/2 life for 'b'
-- buffer = buffer text description
-- delta-g = delta G (delta Gibbs free energy)
-- delta-s = delta S (delta entropy)
-- delta-h = delta H (delta enthalpy)
-- heat-capacity-a = heat capacity of 'a'
-- heat-capacity-b = heat capacity of 'b'
-- other = any other related values (e.g. k1, k2...)
-- source = empirical evidence
-- ********************************************************************************

BIND-kinetics ::= SEQUENCE {
        descr VisibleString OPTIONAL,
        kd RealVal-Units OPTIONAL,
        km RealVal-Units OPTIONAL,
        vmax RealVal-Units OPTIONAL,
        rxn-order RealVal-Units OPTIONAL,
        conc-a RealVal-Units OPTIONAL,
        conc-b RealVal-Units OPTIONAL,
        conc-a-bound RealVal-Units OPTIONAL,
        conc-b-bound RealVal-Units OPTIONAL,
        conc-a-unbound RealVal-Units OPTIONAL,
        conc-b-unbound RealVal-Units OPTIONAL,
        enz-activity-amp-factor RealVal-Units OPTIONAL,
        temp RealVal-Units OPTIONAL,
        ph RealVal-Units OPTIONAL,
        half-life-a RealVal-Units OPTIONAL,
        half-life-b RealVal-Units OPTIONAL,
        buffer VisibleString OPTIONAL,
        delta-g RealVal-Units OPTIONAL,
        delta-s RealVal-Units OPTIONAL,
        delta-h RealVal-Units OPTIONAL,
        heat-capacity-a RealVal-Units OPTIONAL,
        heat-capacity-b RealVal-Units OPTIONAL,
        other SEQUENCE OF BIND-kinetics-other OPTIONAL,
        source BIND-pub-set OPTIONAL
        }

BIND-kinetics-other ::= SEQUENCE {
        descr VisibleString,
        value RealVal-Units
        }


-- ********************************************************************************
-- Dependency of interaction on an experimental condition
--
-- The experimental condition(s) used to observe this chemical action.
-- Pointer to a BIND-condition object.  Uniquely locates an experimental
-- condition by Interaction-id then by Internal-condition-id.
--
```

```
-- Field description for BIND-condition-dependency
-- ***********************************************
-- from-iid = interaction that contains the experimental condition
-- cond = internal condition ID number of the condition description
-- ***************************************************************************

BIND-condition-dependency ::= SEQUENCE {
        from-iid Interaction-id,
        cond Internal-conditions-id
        }

-- ***********************************************************************
-- * Interaction - chemical state for 'a' and/or 'b' (in BIND-descr) *
-- ***********************************************************************

-- ***************************************************************************
-- Chemical state and required chemical state for objects 'a' and 'b'
--
-- The chemical state in the BIND-state-descr is "the chemistry" of 'a' or 'b'
-- in this particular molecular interaction. The chemistry is referred to by
-- reference, typically to another interaction record's
-- interaction:action:result which encodes a BIND-object that is the
-- "bio-processed" form of 'a' or 'b' used in this interaction.
--
-- Field description for BIND-state-descr
-- **************************************
-- a = list of possible chemical states for 'a' that can undergo this
--    interaction
-- a-required-state = the state that 'a' in the above list of possible states
--            is required to assume before interaction takes place.
-- b = list of possible chemical states for 'b' that can undergo this
--    interaction
-- b-required-state = the state that 'b' in the above list of possible states
--            is required to assume before interaction takes place.
-- NOTE: multiple required states are only used if a or b is a molecular complex
--       and the state of more than one sub-unit needs to be described.
--
-- ***************************************************************************

BIND-state-descr ::= SEQUENCE {
        a BIND-state-set OPTIONAL,
        a-required-state SEQUENCE OF BIND-required-state OPTIONAL,
        b BIND-state-set OPTIONAL,
        b-required-state SEQUENCE OF BIND-required-state OPTIONAL
        }

-- ***************************************************************************
-- A set of chemical states
--
-- e.g. multiple phosphorylations on a protein; all of which may be active
--    in this interaction record.
```

```
--
-- Field description for BIND-state-set
-- ************************************
-- max-isid = highest Internal-state-id used in this set
-- states = list of possible chemical states
-- ********************************************************************************

BIND-state-set ::= SEQUENCE {
        max-isid Internal-state-id,
        states SEQUENCE OF BIND-state
        }

Internal-state-id ::= INTEGER


-- ******************************************
-- Interaction chemical state (in BIND-descr)
-- ******************************************
--
--
-- ********************************************************************************
-- A chemical state
--
-- Points to the chemistry of a molecule, if known, by reference to an
-- interaction:action with an explicit 'result' field.
-- This allows conversion of a sequence to chemistry with modifications -
-- can describe a protein that has been phosphorylated at a certain residue,
--
-- Here we can exactly state the chemistry of a molecule as it is found in
-- the cell, even though the top BIND-object may only refer to the GI.
--
-- Field description for BIND-state
-- ********************************
-- isid = Internal-state-id (unique for each state in a BIND-state-set)
-- activity-level = general activity of molecule
-- cause = sequence of actions from this or other Interactions that bring
--          about this state
-- descr = text description (e.g. method used to determine this state)
-- source = empirical evidence for this state
-- sub-unit = if a or b is a molecular complex, specifies the sub-unit to which
--          this state applies
--
-- ********************************************************************************


BIND-state ::= SEQUENCE {
        isid Internal-state-id,
        activity-level ENUMERATED {
                not-specified (0),
                inactive (1),
                very-low (2),
                low (3),
                medium (4),
                medium-high (5),
                high (6),
                very-high (7),
```

```
                extreme (8),
                active (9),
                other (255)
                },
        cause SEQUENCE OF BIND-state-cause OPTIONAL,
        descr VisibleString OPTIONAL,
        source BIND-pub-set OPTIONAL,
        sub-unit BIND-complex-subunit OPTIONAL
        }

-- ****************************************************************************
-- Cause of a chemical state
--
-- The chemical action from this or other interactions that directly brings
-- about this state.
-- References an external interaction:action uniquely.
-- The "cause" is really the Interaction:action pair elsewhere
-- in the database that is the most recent step in the biochemical
-- conversion that forms the biochemical entity in 'a' or 'b'.
-- Action and state are peer BIND-descr tags, this allows
-- a reference to causal 'action' within the chemical state.
--
-- Field description for BIND-state-cause
-- ***************************************
-- from-iid = interaction that contains the causal chemical action
-- cause = internal action ID number that caused this activity
-- ****************************************************************************

BIND-state-cause ::= SEQUENCE {
        from-iid Interaction-id,
        cause Internal-action-id
        }

-- ****************************************************************************
-- A required chemical state for interaction to take place
--
-- The state in the state set that is required for the interaction to take
-- place. Uniquely locates a chemical state within this interaction record
-- by Internal-state-id.
--
-- Field description for BIND-required-state
-- ***************************************
-- isid = Internal-state-id of the required state.  Points to
--          one chemical state in the BIND-state-set in the same record
-- descr = description of state requirement
-- source = empirical evidence
-- ****************************************************************************

BIND-required-state ::= SEQUENCE {
        isid Internal-state-id,
        descr VisibleString OPTIONAL,
        source BIND-pub-set OPTIONAL
        }
```

```
-- <!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!>
-- ********************
-- * Molecular-Complex *
-- ********************

-- ****************************************************************************
-- A set of Molecular Complexes
--
-- Field description for BIND-Complex-set
-- **************************************
-- date = date this set of records was collected
-- database = name and description of database that this set comes from
-- complexes = set of molecular complex records
-- ****************************************************************************


BIND-Complex-set ::= SEQUENCE {
        date Date OPTIONAL,
        database BIND-Database-site OPTIONAL,
        complexes SEQUENCE OF BIND-Molecular-Complex
        }


-- ****************************************************************************
-- A molecular complex record
--
-- A collection of more than two interactions that form a complex.
-- i.e. Three or more BIND-objects that operate as a unit.  It is a
-- useful shorthand when defining BIND pathways.
--
-- A molecular complex can also be defined if the interactions in it are not
-- completely known. Create interactions with molecule 'a' as the sub-unit of
-- the complex and molecule 'b' as 'not-specified' for all of the known
-- sub-units.
--
-- Field description for BIND-Molecular-Complex
-- ********************************************
-- date = date of record entry
-- updates = a list of updates for the record
-- mcid = molecular complex accession number.
-- descr = text description of complex (e.g. ribosome)
-- sub-num = total number of sub-units in this complex
-- sub-units = collection of BIND-objects in the complex
-- interaction-list = list of interactions in this complex
-- ordered = TRUE if order of interactions is known and
--              interaction-list is ordered in this way
-- complex-topology = a connectivity graph of the complex topology
--                      with BIND-objects as nodes
-- exclusive-int = interactions that exclusively occur between subunits only
--              when part of this molecular complex
-- source = empirical evidence references
-- authors = person(s) who authored this record.
-- priv = TRUE if this complex is private
-- ****************************************************************************
```

```
BIND-Molecular-Complex ::= SEQUENCE {
        date Date,
        updates SEQUENCE OF BIND-update-object OPTIONAL,
        mcid Molecular-Complex-id,
        descr VisibleString OPTIONAL,
        sub-num BIND-mol-sub-num,
        sub-units SEQUENCE OF BIND-mol-object,
        interaction-list SEQUENCE OF Interaction-id,
        ordered BOOLEAN DEFAULT FALSE,
        complex-topology SEQUENCE OF BIND-mol-object-pair OPTIONAL,
        exclusive-int SEQUENCE OF BIND-Interaction OPTIONAL,
        source BIND-pub-set,
        authors SEQUENCE OF Author OPTIONAL,
        priv BOOLEAN DEFAULT FALSE
        }

Molecular-Complex-id ::= INTEGER

-- specify a specific sub-unit of a complex
BIND-complex-subunit ::= SEQUENCE {
        mcid Molecular-Complex-id,
        bmoid BIND-mol-object-id    =
        }


-- ************************************************************************
-- Sub unit numbers in a Molecular Complex
--
-- This number can be an integer or a fuzzy integer.
--
-- Field description for BIND-Complex-set
-- ************************************
-- num = integer number of sub-units
-- num-fuzz = fuzzy integer number of sub-units (e.g. microtubule, virus)
-- ************************************************************************


BIND-mol-sub-num ::= CHOICE {
        num INTEGER,
        num-fuzz Int-fuzz
        }


-- ************************************************************************
-- A graph describing topology of a molecular complex
--      BIND-mol-object objects are nodes in the graph
--      BIND-mol-object-pair objects are edges in the graph
--
-- Field description for BIND-mol-object
-- ************************************
-- bmoid = internal ID BIND-object
-- sub-unit = a sub-unit in a molecular complex
-- num = number of this sub-unit
--
-- Field description for BIND-mol-object-pair
-- ******************************************
```

```
-- a-bmoid = the sub-unit pointed to by this ID is connected to...
-- b-bmoid = the sub-unit pointed to by this ID
-- ****************************************************************************


BIND-mol-object ::= SEQUENCE {
          bmoid BIND-mol-object-id,
          sub-unit BIND-object,
          num BIND-mol-sub-num OPTIONAL
          }

BIND-mol-object-pair ::= SEQUENCE {
          a-bmoid BIND-mol-object-id,
          b-bmoid BIND-mol-object-id
          }

BIND-mol-object-id ::= INTEGER


-- <!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!>
-- *********************************
-- * Biomolecular chemical pathway *
-- *********************************

-- ****************************************************************************
-- A set of Pathways
--
-- Field description for BIND-Pathway-set
-- ***************************************
-- date = date this set of records was collected
-- database = name and description of database that this set comes from
-- pathways = set of pathway records
-- ****************************************************************************


BIND-Pathway-set ::= SEQUENCE {
          date Date OPTIONAL,
          database BIND-Database-site OPTIONAL,
          pathways SEQUENCE OF BIND-Pathway
          }


-- ****************************************************************************
-- A pathway record.
--
-- A collection of more than two interactions that form a pathway.
-- i.e. Three or more BIND-objects that are generally free from each
-- other, but form a network of interactions.
--
-- Field description for BIND-pathway
-- **********************************
-- date = date of record entry
-- updates = a list of updates for the record
-- pid = pathway accession number
-- pathway = a collection of interactions and signal modification objects
-- descr = description of a pathway
```

```
-- source = empirical evidence references
-- authors = person(s) who authored this record.
-- priv = TRUE if this pathway is private
-- ********************************************************************************

BIND-Pathway ::= SEQUENCE {
        date Date,
        updates SEQUENCE OF BIND-update-object OPTIONAL,
        pid Pathway-id,
        pathway SEQUENCE OF Interaction-id,
        descr BIND-path-descr,
        source BIND-pub-set,
        authors SEQUENCE OF Author OPTIONAL,
        priv BOOLEAN DEFAULT FALSE
        }

Pathway-id ::= INTEGER


-- ********************************************************************************
-- Pathway description
--
-- Field description for BIND-path-descr
-- ************************************
-- descr = text description of pathway
--         (e.g. lipid biosynthesis, bacterial chemotaxis, Ras pathway, etc.)
-- cell-cycle = stage of a cell cycle that this pathway is in effect
-- pathological-state = disease manifestation if this pathway is present
-- pathway-actions = list of chemical actions that occur in the pathway
-- ********************************************************************************


BIND-path-descr ::= SEQUENCE {
        descr VisibleString OPTIONAL,
        cell-cycle SEQUENCE OF BIND-cellstage OPTIONAL,
        pathological-state SEQUENCE OF BIND-pathol-state OPTIONAL,
        pathway-actions SEQUENCE OF BIND-state-cause OPTIONAL
        }

BIND-pathol-state ::= SEQUENCE {
        pathway-iid Interaction-id,
        interaction CHOICE {
                ablated NULL,
                replaced-by Interaction-id
        },
        pathol-state VisibleString,
        descr VisibleString OPTIONAL,
        source BIND-pub-set
        }


-- <!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!><!>
-- ********************************
-- * BIND Cross Reference System *
-- ********************************
```

```
-- ********************************
-- * IID cross reference record *
-- ********************************

-- ****************************************************************************
-- Cross reference for IID searching
--
-- Field description for BIND-Iid-Cross-Ref
-- ****************************************
-- iid = Interaction ID
-- iids = list of interactions that contain this iid
-- pids = list of pathways that contain this iid
-- mcids = list of molecular complexes that contain this iid
-- ****************************************************************************


BIND-Iid-Cross-Ref ::= SEQUENCE {
        iid Interaction-id,
        iids SEQUENCE OF Interaction-id OPTIONAL,
        pids SEQUENCE OF Pathway-id OPTIONAL,
        mcids SEQUENCE OF Molecular-Complex-id OPTIONAL
        }


-- ********************************
-- * MCID cross reference record *
-- ********************************

-- ****************************************************************************
-- Cross reference for MCID searching
--
-- Field description for BIND-Mcid-Cross-Ref
-- ****************************************
-- mcid = Molecular Complex ID
-- iids = list of interactions that contain this mcid
-- pids = list of pathways that contain this mcid
-- mcids = list of molecular complexes that contain this mcid
-- ****************************************************************************


BIND-Mcid-Cross-Ref ::= SEQUENCE {
        mcid Molecular-Complex-id,
        iids SEQUENCE OF Interaction-id OPTIONAL,
        pids SEQUENCE OF Pathway-id OPTIONAL,
        mcids SEQUENCE OF Molecular-Complex-id OPTIONAL
        }


-- ********************************
-- * GI/DI cross reference record *
-- ********************************

-- ****************************************************************************
-- Cross reference for gi/di searching
--
-- Field description for BIND-Cross-Ref
```

```
-- ************************************
-- gi = gi number
-- di = di number
-- iids = list of interactions that this gi is involved in
-- pids = list of pathways that this gi is involved in
-- mcids = list of molecular complexes that this gi is involved in
-- ********************************************************************************


BIND-Cross-Ref ::= SEQUENCE {
        gi Geninfo-id DEFAULT 0,
        di Domain-id DEFAULT 0,
        iids SEQUENCE OF Interaction-id,
        pids SEQUENCE OF Pathway-id OPTIONAL,
        mcids SEQUENCE OF Molecular-Complex-id OPTIONAL
        }


-- ************************************
-- * PMID/MUID cross reference record *
-- ************************************
                                    =
-- ********************************************************************************
-- Cross reference for pmid/muid searching
--
-- Field description for BIND-Pub-Cross-Ref
-- ****************************************
-- uid = muid or pmid
-- iids = list of interactions with this publication as a reference
-- pids = list of pathways with this publication as a reference
-- mcids = list of molecular complexes with this publication as a reference
-- ********************************************************************************


BIND-Pub-Cross-Ref ::= SEQUENCE {
        uid INTEGER,
        iids SEQUENCE OF Interaction-id,
        pids SEQUENCE OF Pathway-id OPTIONAL,
        mcids SEQUENCE OF Molecular-Complex-id OPTIONAL
        }


END
39


69
```

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic.Acids.Res.*, **25**, 3389-3402.

Bairoch, A. & Apweiler, R. (1999) The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic.Acids.Res.*, **27**, 49-54.

Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J., Ouellette, B.F., Rapp, B.A. & Wheeler, D.L. (1999) GenBank. *Nucleic.Acids.Res.*, **27**, 12-17.

Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer, E.F.J., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1978) The protein data bank: a computer-based archival file for macromolecular structures. *Arch.Biochem.Biophys.*, **185**, 584-591.

DDBJ/EMBL/GenBank. (1997) *The DDBJ/EMBL/GenBank Feature Table Definition Version 2.1*

Eilbeck, K., Brass, A., Paton, N. & Hodgman, C. (1999) INTERACT: an object oriented protein-protein interaction database. *Ismb.*, **7**, 87-94.

Gasteiger, J. (1996) Chemical Information in 3D-Space. *J.Chem.Inf.Comput.Sci.*, **36**, 1030-1037.

Higgins, D.G., Thompson, J.D. & Gibson, T.J. (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.*, **266**, 383-402.

Hogue, C.W., Ohkawa, H. & Bryant, S.H. (1996) A dynamic look at structures: WWW-Entrez and the Molecular Modeling Database. *Trends.Biochem.Sci.*, **21**, 226-229.

Igarashi, T. & Kaminuma, T. (1997) Development of a cell signaling networks database. *Pac.Symp.Biocomput.*, 187-197.

Kans, J.A. & Ouellette, B.F. (1998) Submitting DNA Sequences to the Databases . In Baxevanis, A.D. & Ouellette, B.F. (eds), Bioinformatics. John Wiley & Sons Toronto pp.319-353.

Marcotte, E.M., Pellegrini, M., Ng, H.L., Rice, D.W., Yeates, T.O. & Eisenberg, D. (1999) Detecting protein function and protein-protein interactions from genome sequences. *Science*, **285**, 751-753.

Mendelsohn, A.R. & Brent, R. (1999) Protein interaction methods--toward an endgame. *Science*, **284**, 1948-1950.

Mohr, E., Horn, F., Janody, F., Sanchez, C., Pillet, V., Bellon, B., Roder, L. & Jacq, B. (1998) FlyNets and GIF-DB, two internet databases for molecular interactions in Drosophila melanogaster. *Nucleic.Acids.Res.*, **26**, 89-93.

Object Management Group. (1996) *CORBA Architecture and Specifications* , OMG Publications

Ostell, J. & Kans, J.A. (1998) The NCBI Data Model . In Baxevanis, A.D. & Ouellette, B.F. (eds), Bioinformatics. John Wiley & Sons pp.121-144.

Schuler, G.D., Epstein, J.A., Ohkawa, H. & Kans, J.A. (1996) Entrez: molecular biology database and retrieval system. *Methods Enzymol.*, **266**, 141-162.

Stoesser, G., Tuli, M.A., Lopez, R. & Sterk, P. (1999) The EMBL Nucleotide Sequence Database. *Nucleic.Acids.Res.*, **27**, 18-24.

Sugawara, H., Miyazaki, S., Gojobori, T. & Tateno, Y. (1999) DNA Data Bank of Japan dealing with large-scale data submission. *Nucleic.Acids.Res.*, **27**, 25-28.

Weininger, D. (1988) SMILES, a Chemical Language and Information System. *J.Chem.Inf.Comput.Sci.*, **28**, 31-36.

5